TREATMENT OF PREMATURE OVARIAN FAILURE

Cartwright, Beth Rhian

Awarding institution:
King's College London

TREATMENT OF PREMATURE OVARIAN FAILURE

A prospective open randomised controlled trial to compare the effects of active treatment with hormone replacement therapy or the combined oral contraceptive pill and observation of patients who choose to have no treatment, on bone density and other parameters over two years.

BETH RHIAN CARTWRIGHT

MBBS MRCOG

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF MEDICINE, KING'S COLLEGE LONDON, 2014

Abstract

Hormone replacement therapy (HRT) and the combined oral contraceptive pill (COCP) are both commonly used for oestrogen replacement in premature ovarian failure but there is a lack of evidence of their effects. We compared the effects of combined HRT (Nuvelle) and the COCP (Microgynon 30) in 30 women with premature ovarian failure in a two year open-label randomised trial. 29 women who declined to take treatment were also followed using the same protocol. 36 women (61%) completed the trial (15 in the no treatment group; 12 in the HRT group; 9 in the COCP group). In comparison with the COCP, treatment with HRT increased bone mineral density at the lumbar spine at 2 years, which was the primary outcome measure (+0.038 g/cm^2; 95% CI 0.002 to 0.073; p 0.040; linear regression using adjustment for baseline values). Bone turnover markers (P1NP and CTX) showed similar reductions in the two treatment groups. There were trends in favour of HRT in lipid profile, high sensitivity C-reactive protein, blood pressure and sexual function. The HRT group had a significantly greater reduction in menopausal symptoms at 24 months. Improvements in most symptom scores took longer in the COCP group.

In the no treatment group, bone density dropped at all sites over the course of the trial. The no treatment group also performed poorly in comparison with the treatment groups in bone turnover, depression score and menopausal symptoms score.

These findings will have important implications for counselling young women with premature ovarian failure on their choice of oestrogen replacement. We have shown that in many respects HRT performed superiorly to the COCP. However, further research is required to confirm these effects. The results from the no treatment group will enable women who choose to decline treatment to be counselled on the effects of this choice of management.

Table of contents

List of Figures

List of Tables

Acknowledgements

The work submitted in this thesis was carried out in the HRT Research Unit at Guy's Hospital, London. I am hugely grateful to my supervisor Janice Rymer who has always been incredibly encouraging and supportive. I would like to thank both her and my second supervisor Ignac Fogelman for reading and commenting on this thesis at its various stages. I would also like to express my gratitude to Jill Robinson who gave me a great deal of advice on setting up and running the project.

I would like to thank my mum who looked after my first baby when he was tiny and I needed to carry on seeing the trial participants. I also thank my partner Jeremy for lots of patience throughout the project and whilst I have been writing up, and for looking after the children on numerous occasions so that I could finish it!

I am very grateful to the Rosetrees Trust and the Moulton Foundation who provided funding for the consumables for the study.

I am grateful to Paul Seed for his statistical advice on both the sample size and data analysis. I would like to thank the radiographers in the Osteoporosis Unit at Guy's Hospital, under the direction of Christine O'Gara, who performed the bone scans and were incredibly accommodating with time slots when the participants needed to change their appointments. I am grateful to Zehra Arkir and Darren Costello from GSTS Pathology who oversaw the storage and analysis of the blood samples.

The trial would not have been possible without the dedication of the participants, some of whom travelled very large distances to take part in the trial, and I would like to thank them for their commitment to the project.

Chapter 1 Introduction and Aims of Thesis

1.1 Introduction

Premature ovarian failure (POF), sometimes called early menopause, is a devastating condition. It is defined as a loss of ovarian function before the age of 45 (British Menopause Society definition) (1). POF is not uncommon, spontaneously affecting 1% of women under the age of 40 and 5% before 45 (2). POF is not the same as the normal menopause (which occurs at an average age of 52 in the UK) because in POF intermittent ovarian function may continue in an unpredictable fashion for many years. This has been demonstrated by ultrasound studies following the growth of ovarian follicles and the observation of spontaneous pregnancies occurring after the diagnosis of POF in 5-10% of women (3, 4).

The diagnosis of POF carries enormous impact for a young woman from both a physical and psychological perspective. It will affect almost every aspect of her life, from causing menopausal symptoms to influencing long term health risks, as well as meaning that there is little chance of conceiving a pregnancy naturally.

It is worth noting that some countries use the cut-off of 40 rather than 45 to define POF. However, a woman who experiences POF at the age of 44 has 8 extra years of oestrogen deficiency compared with a woman who has her menopause at 52. She is exposed to all the risks of early oestrogen deficiency, including bone loss. We therefore use the British Menopause Society's definition both in clinical practice and for this research project.

In POF the ovaries stop producing normal levels of oestrogen. The resulting low oestrogen level has many consequences. Arguably the most important is its effect on the bone, where a lack of oestrogen causes early bone loss which can ultimately lead to osteoporosis. Osteoporosis increases the likelihood of bone fractures, which can lead to considerable morbidity and decreased life expectancy – for example data suggest a death rate of 20-30% from all causes within one year following a fractured neck of femur (5). A recent service evaluation of women seen in the POF clinic at GSTT (Guy's and St Thomas' NHS Foundation Trust) revealed that loss of bone density is the main concern for over 70%. Previous trials investigating bone density in POF have been observational and retrospective. Several studies have shown that women with POF have reduced bone density, and one retrospective study has shown that they have a lifelong increased risk of fractures (6-10). To date, there have been no prospective studies in spontaneous POF investigating the quantitative effects on bone density of different treatments or no treatment.

Another long-term consequence of POF is increased cardiovascular risk (11, 12) This may in part be due to a change in lipid profile seen in POF (13). Osteoporosis and cardiovascular events can seriously affect life expectancy and quality of life. Cardiovascular disease is the leading cause of mortality in women. It is likely that the management of POF at the time of diagnosis will affect the development and impact of osteoporosis and cardiovascular disease in older age.

Studies also indicate that women with POF have high levels of depression and sexual dysfunction (14, 15). However, no previous studies have assessed the full range of symptoms that may be experienced in POF, or the effects of different treatments or no treatment on these.

Oestrogen replacement is recommended for women with POF, either in the form of hormone replacement therapy or the combined oral contraceptive pill, in order to alleviate menopausal symptoms and protect against the long term health sequelae of bone loss and increased cardiovascular risk. However, the most suitable form of oestrogen replacement is unknown and management is not currently evidence-based. This is acknowledged in most reviews on the subject.(1, 16, 17) In 2004, the RCOG Menopause and Hormone Replacement study group advised that research is recommended 'to develop and assess treatment strategies in women with premature menopause'. Very few studies have as yet been carried out. Hormone replacement therapy or the combined oral contraceptive pill are the two broad choices but there is very little evidence-based information on their relative advantages and disadvantages in this context. This makes it extremely difficult to provide accurate advice to affected women.

HRT is considered to be more 'physiological' but some young women dislike taking a preparation which they associate with older women. Some women may also find that a standard dose of HRT is not adequate to control their symptoms. The COCP contains synthetic oestrogen at a higher dose but with a 'pill free' week. It may be the preferred choice for some young women as it is a medication they are familiar with, it does not carry HRT's association with older women, and it is generally more 'peer-friendly'. Women with POF face many years of oestrogen replacement and deserve to know more about the relative advantages and disadvantages of different treatments, particularly in relation to long term health outcomes such as preventing bone loss and minimising cardiovascular risk.

There are also a significant proportion of women who decline treatment as they feel it is 'unnatural' or are concerned about the safety or side effects. There have been no prospective studies in spontaneous POF investigating the effects of different treatments or no treatment. In a patient-centred approach to health care, these data are vital to enable affected women to make treatment choices.

This study assesses the effects of hormone replacement therapy, the combined oral contraceptive pill and no treatment on bone density and turnover, cardiovascular risk, ovarian function, menopausal symptoms, sexual function, depression and quality of life in women with premature ovarian failure. The results will help women make informed decisions about their method of treatment.

The term 'premature ovarian failure' is being succeeded by the term 'premature ovarian insufficiency' or even 'primary premature ovarian insufficiency'. These new terms are more accurate, in that ovarian function does not necessarily fail completely and permanently, and they are considered to be more sensitive for affected women. However, the term 'premature ovarian failure' has been used throughout this thesis because the trial was named and started before these new terms became widely used in the United Kingdom.

1.2 Aims of Thesis

The aim of this study was to compare the effects of a combined hormone replacement therapy (Nuvelle; oestradiol 2mg daily, plus levonorgestrel 75mcg for 12 days a month) with a combined oral contraceptive pill (Microgynon 30; ethinyloestradiol 30mcg and levonorgestrel 150mcg taken for 3 weeks then with a one week break) on bone density and other parameters in women with premature ovarian failure, and to observe the same outcomes in women who chose not to take any treatment.

We hypothesised that in the two treatment groups (HRT and the COCP) bone density would be maintained and that women in the no treatment group would experience a decrease in bone density at the normal post-menopausal rate (2-3% per year initially and approximately 1% per year subsequently). However, due to the intermittent nature of POF we also considered the possibility that in some untreated women bone loss would be less severe.

The primary outcome was a change in lumbar spine bone mineral density (assessed by dual energy X-ray absorptiometry; DXA) at 24 months.

Secondary outcomes were changes in:

- Total hip and femoral neck bone mineral density
    - assessed by DXA
- Cardiovascular markers
    - assessed by total cholesterol, LDL cholesterol, triglycerides, high sensitivity C-reactive protein
- Ovarian function
    - assessed by trans-vaginal ultrasound for antral follicle count and ovarian volume and serum anti-Mullerian hormone and inhibin B
- Menopausal symptoms
    - assessed by Greene Climacteric Scale and Menopause Symptoms Treatment Satisfaction Questionnaire
- Sexual Function
    - assessed by Brief Profile of Female Sexual Function
- Depression
    - assessed by Patient Health Questionnaire-9
- Quality of Life
    - assessed by Short Form-36

Chapter 2 Literature Review

2.1 Pathophysiology of premature ovarian failure (POF)

Premature ovarian failure (POF) is defined as a loss of ovarian function before the age of 45 (British Menopause Society) (1). POF can be secondary to chemotherapy, radiotherapy, pelvic surgery, infection or removal of the ovaries, or it can happen spontaneously. The prevalence of spontaneous POF before the age of forty is estimated at 1% and before forty-five at 5% (2). In the majority of cases of spontaneous POF the cause is unknown (idiopathic POF). Recognised causes include autoimmune disease, Fragile X premutation and other genetic disorders (table 1).

The mechanism of idiopathic spontaneous POF is unknown. A reduced initial number of oocytes, accelerated oocyte atresia and follicular dysfunction have been suggested. POF is different from the 'normal' menopause because ovarian function may be intermittent and unpredictable. Pregnancy may occur, sometimes many years after diagnosis (3), and studies monitoring ovarian activity have shown that over 80% have ovarian follicular activity and almost 50% ovulate (4).

Causes of spontaneous POF	**Causes of secondary POF**
No cause found (idiopathic)	Bilateral oophorectomy
Autoimmune disease	Chemotherapy
Fragile X premutation	Radiotherapy
Other genetic causes – e.g. Turner's syndrome, Down's syndrome, rare mutations such as FSH receptor polymorphism and inhibin B mutation	Interruption of ovarian blood supply due to pelvic surgery or uterine artery embolisation
Enzyme deficiencies – e.g. galactosaemia,	Infections – mumps, tuberculosis, HIV

Table 1 Causes of spontaneous and secondary premature ovarian failure

2.1.1 Mechanism of normal menopause

The median age of menopause in the UK is 52 (18). The ovaries have a finite number of oocytes and a certain number undergo atresia each month, starting antenatally. The rate of decline of ovarian follicles in humans is more complicated than the simple exponential decline seen for example in rats. Using histological data from several studies estimating ovarian follicle number at different ages, Faddy (2000) proposes a model for follicle decline where the rate of

change of number of follicles is proportional to the number remaining. This means that there is little change in the rate of exponential decline until the late 30s, when the number of follicles is approximately 25,000, and the rate of decline increases sharply (19). This model was found to correspond very closely with the distribution of age at menopause, with the timing of menopause corresponding with a median pool of 1100 follicles (19). Another model using new histological data also estimated that at the menopause about 1000 follicles remain (20).

As the number of follicles approaches this threshold, the ovaries become less responsive to follicle stimulating hormone (FSH) produced by the pituitary, less follicles develop and oestradiol levels decrease. The loss of negative feedback of oestradiol and other ovarian hormones causes serum FSH to rise. In the early peri-menopause the remaining follicles occasionally respond to the high FSH levels and start to develop and produce oestradiol, causing significant fluctuations in serum oestradiol and FSH. This stops as the follicles become more depleted and eventually the FSH remains high, the ovaries become completely inactive and serum oestradiol becomes very low.

2.1.2 Spontaneous premature ovarian failure

A cause for spontaneous POF is identified in approximately 35%, although the sampling bias in studies of women attending clinics could render this an over-estimate. A genetic aetiology is reported in approximately 5% (21) and autoimmune in up to 30% (22), although this is likely to depend heavily on the clinic referral base. At least 65% have no identifiable cause and it is likely that this figure is much higher for the general POF population. At Guy's and St Thomas' 85% with spontaneous POF had no cause identified from routine clinic investigations (23). The largest study investigating causes of POF reviewed 357 women aged under 40 years at diagnosis who were referred to the national POF centre in France (24). 14% were found to have signs of clinical or biochemical autoimmunity and 7% had a genetic cause. Of note, 24% of the women in this study presented with primary amenorrhoea; unfortunately the authors do not provide separate data for this group but report that in the majority there was no identifiable aetiology.

2.1.2.1 Genetic causes

Many cases of spontaneous POF appear to be inherited, although estimates have varied widely, from approximately 4 to 31% (25). Whilst causative mutations in some genes have been identified (26), it is not practical to test for these routinely and it is likely that there are many more which are as yet undiscovered. The two genetic causes of POF which can be readily tested for are the FMR1 (fragile X) premutation and chromosomal disorders, such as Turner syndrome, which can be assessed on karyotype.

2.1.2.1.1 Fragile X premutation

The FMR1 (fragile X) premutation, defined as 55 to 200 CGG repeats in the FMR1 gene, is responsible for approximately 4-5% of cases of POF overall (21, 27). In women with a positive family history of POF, approximately 10% test positive for the premutation, compared with 2%

with no family history (27). It is important to identify those carrying the FMR1 premutation for several reasons. In the event of spontaneous conception these individuals are at risk of having a child with fragile X syndrome; 3 such cases have been reported (28). For medico-legal reasons it is important to inform women of this possibility. Other issues to consider are disclosure and offering testing to other family members and that the premutation can occasionally be associated with a neurodegenerative disorder (29). The premutation has been estimated to occur in between 1 in 113 to 1 in 259 women in the general population (30-32). Women with the FMR1 premutation have a risk of approximately 15% of developing POF (33).

The FMR1 gene codes for a RNA binding protein called fragile X mental retardation protein (FMRP). In the full mutation, no FMRP is produced. In the premutation, there is a normal amount of FMRP but excess messenger RNA production, which may be the cause of ovarian damage. However, the exact mechanism behind this is unknown.

2.1.2.1.2 X chromosome disorders

The X chromosome contains genes which are vital for normal ovarian development and function. Turner syndrome, characterised by the absence of all or part of the second X chromosome, affects approximately 1 in 2500 live female babies (34). The number of primordial follicles in utero is reduced and these then undergo premature apoptosis, resulting in ovarian dysgenesis and streak ovaries. The mechanisms behind this are unknown. Most girls with a 45 XO karyotype do not enter puberty but 40% of those with 45X/46XX mosaicism do; this is usually followed by premature ovarian failure (34). This area is incompletely understood. There are cases of spontaneous pregnancy in women with non-mosaic Turner's syndrome, indicating a presence of mature ovarian follicles in some women with this genotype (35). There has been much interest recently in cryopreservation of oocytes or ovarian tissue in these women whose fertility may be reduced in later life (36).

A recent study investigated low levels of X-monosomy mosaicism in 1000 young women (mean age 25) with spontaneous POF and no clinical signs of Turner's syndrome using karyotype and FISH (interphase flourescent in situ hybridisation) techniques. It found chromosome abnormality rates of 14% in women with primary amenorrhoea and 8% with secondary amenorrhoea using karyotype alone. The FISH technique was used to detect low levels of 45XO/46XX mosaicism and found that 11.5% of POF patients with a normal karyotype had an increased rate of mosaicism (37).

2.1.2.1.3 X chromosome gene defects

As well as a complete lack of part of an X chromosome, gene defects can also cause ovarian failure. Two regions on the X chromosome named POF1 and POF2 have been identified as necessary for normal ovarian function. POF1 is found at Xq13-Xq26 and is of maternal origin. POF2 is found at Xq13.3-q21.1 and is of paternal origin (38). Disruption of these genes, including deletion and translocation, is associated with POF.

The bone morphogenic protein-15 (BMP15) is a protein coded on the long arm of the X chromosome (Xp11.2) which is expressed in the oocyte during folliculogenesis. It has been found in vitro and in animal models to have a positive effect on folliculogenesis. Studies of the BMP15 gene in the POF population have found mutations in up to 10% (39).

2.1.2.1.4 Autosomal mutations

Autosomal mutations can also be associated with POF. These include mutations of the phosphomannomutase 2 gene, the FSH receptor gene, the galactose-1-phosphate uridyltransferase gene, the blepharophimosis gene, and the autoimmune regulator gene. Several of these mutations cause enzyme deficiencies which are serious or potentially lethal – for example carbohydrate-deficient glycoprotein syndrome type 1 caused by a lack of phosphomannomutase. A lack of galactose-1-phosphate uridyltransferase causes galactosaemia, which causes a reduction in germ cell development during fetal oogenesis, presumably due to the toxic effect of galactose (40). Mutations in the FOXL2 gene on chromosome 3 cause blepharophimosis ptosis epicanthus inversus syndrome (BPES). BPES type 1 but not BPES type 2 is associated with POF. The autoimmune regulator gene is responsible for autoimmune polyendocrinopathy candidiasis actodermal dystrophy, of which POF can be a part.

Inhibin is produced by granulosa cells and acts to reduce FSH secretion by the pituitary in a negative feed-back fashion. It also has actions in the ovary itself, including stimulation of androgen synthesis in thecal cells and regulation of follicle growth. These roles led to it being considered as a candidate gene for POF and mutations in the inhibin gene have been linked to POF in some populations (41).

2.1.2.2 Autoimmune cause

30% of cases of POF are estimated to be due to autoimmune disease (22). However, this may vary depending on the age of patients seen and referral patterns; at Guy's and St Thomas' only 3% have associated autoimmune disease (23). There is no value in assessing anti-ovarian antibodies because these are also present in a large percentage of the general population (42). However, a positive anti-thyroid antibody result, which is found in approximately 24% of women with POF (43), and a family history of auto-immune disease, can be used to identify a presumed autoimmune aetiology. It is also worthwhile testing for anti-adrenal antibodies to identify the small subgroup of women with POF who are at risk of developing Addison's disease (44, 45). This is especially important if egg donation is being pursued because of the danger of cortisol deficiency in pregnancy.

2.1.2.3 Idiopathic premature ovarian failure

The mechanism of most cases of spontaneous POF is unknown. A reduced initial number of oocytes, accelerated oocyte atresia and follicular dysfunction have been suggested. POF is different from the 'normal' menopause because ovarian function may be intermittent and

unpredictable. Pregnancy may occur, sometimes many years after diagnosis (3), and studies monitoring ovarian activity have shown that over 80% have ovarian follicular activity and almost 50% ovulate (4).

Histological data has shown a large variation in follicle numbers at different ages. Due to the nature of the rate of follicular decline, it appears that the number of initial oocytes would need to be very significantly reduced to cause POF. Using their model for follicular depletion, Faddy et al estimate that a loss of 90% of follicles before the age of 14 would result in the menopausal threshold of 1000 follicles at the age of 27 (46). However, studies have shown that in women who go through the menopause at the usual time very few if any follicles are left post-menopause (47) and this theory would not explain cases where pregnancy has occurred in women with POF many years after the diagnosis.

2.1.3 Secondary premature ovarian failure

Premature ovarian failure can occur following treatment with certain types of chemotherapy or radiotherapy which damage ovarian tissue. Abdominal radiotherapy and the protocols required prior to bone marrow transplantation are associated with high rates of permanent ovarian failure. Alkylating agents have a detrimental effect on ovarian function in the longer term (48). Ovarian function post treatment appears to be more favourable in pre-pubertal than post-pubertal girls (48). Methods used to reduce rates of ovarian damage include concurrent treatment with gonadotrophin releasing hormone agonists, the mechanism of action of which is unknown, and transposition of the ovaries. Cryopreservation of oocytes, embryos or ovarian tissue may also be options to preserve future fertility.

Infections can also cause ovarian failure. Mumps has long been recognised as a potential cause, although the ovarian failure may be transient (38, 49). HIV infection is a relatively newly recognised cause (50). Ovarian tuberculosis is also described as a cause of POF (38).

Other treatments which may decrease ovarian reserve include those which have the potential to affect blood supply to the ovaries, such as hysterectomy with ovarian conservation, myomectomy and uterine artery embolisation. A reduction in ovarian tissue, for example due to endometriosis, or the removal of endometriomas or other ovarian cysts, can also reduce ovarian reserve (51).

Recent data from 4968 women born in a single week in 1958 revealed that POF by the age of 40, either spontaneous or iatrogenic, affected 7.4% of those followed up (52).

2.2 Consequences of POF

A low oestrogen level causes many unpleasant symptoms and in the long term has a detrimental effect on bone and cardiovascular health.

2.2.1 Menopausal symptoms

Menopausal symptoms are reported to affect approximately 70% of climacteric women (38). These include vasomotor symptoms (hot flushes and night sweats), vaginal dryness, low mood, decreased libido, lack of energy and irritability. All of these can impact quality of life. Oestrogen replacement effectively reduces most symptoms, especially vasomotor symptoms (53).

The prevalence of most of the symptoms of oestrogen deficiency in the spontaneous POF population and the effects of treatment has not been studied. It is possible that there are differences in initial symptoms compared with the older group of women due to intermittent ovarian function. Depression and sexual function are the best studied symptoms, but data on the impact of treatments or the natural course of the disorders is lacking.

A questionnaire study conducted in 2000 in a London teaching hospital investigated 'psychological well-being' in 64 women with spontaneous POF and found depression scores comparable to a psychiatric in-patient population (14). 74% of respondents were classified as depressed. Rates of depression decreased with increasing years since diagnosis but were unaffected by whether the woman already had a child or not. Unfortunately there was a low response rate of 42%. A more recent study in America assessed depression and other mood disorders in 174 women with POF using a structured clinical interview (54). They also found high rates of depression in POF (67% of the POF sample had a mood disorder, mostly depression, compared with reported rates of 24% from community based studies). This study also investigated the timing of onset of depression and found that in most women the diagnosis of depression came before that of POF but after menstrual cycle irregularity. This could be explained by the theory of early ovarian insufficiency causing depression. However the authors then compare their data with data collected in their unit from 100 women with Turner syndrome and found significantly lower rates of depression in the Turner syndrome group. This led them to query whether POF and depression could share pathophysiology.

Singer et al (2011) carried out a questionnaire study by post to women attending London clinics and online via the Daisy Network, assessing the psychosocial aspects of POF (55). The response rate was 62% to the postal questionnaire and altogether 136 responses were analysed. Quality of life was assessed using the Short Form-36 (SF-36) questionnaire and found that compared with normative data, women with POF had worse mental and physical health and rated their quality of life as relatively low. A negative impact on self-image and confidence was also reported. The 1958 Birth Cohort study also found that women with both iatrogenic and spontaneous POF had a reduced SF-36 quality of life score (52).

Two studies have been carried out assessing the effect of HRT on symptoms in women with iatrogenic POF. A questionnaire study of 89 women who had early menopause secondary to radiotherapy or chemotherapy indicated that women choosing to take hormone treatment (type not specified) have a lower menopausal symptom score and better physical component of quality of life score (assessed by SF-36), as compared with those not taking treatment (56). However, the women in the treatment group were significantly younger (39 vs 44 years).

Another study of 31 women with chemotherapy-induced POI (mean age 36) investigated the effect of HRT in those women who were happy to take it (15/31) (57). Compared with those who did not take HRT, this group had a significant reduction in hot flushes, insomnia and psychological and emotional changes (all improved in 66%), vulvo-vaginal and skin atrophy (53%), genito-urinary disturbances (53%) and musculskeletal symptoms (53%).

Decreasing stigmatisation, goal readjustment, positive affect and education about the condition have all been identified as factors which can contribute to emotional well-being in POF (58).

Questionnaire studies indicate that women with POF have lower sexual satisfaction and sexual function scores than controls (15, 59). 50% of these women who completed the Short Personal Experiences Questionnaire in the Singer et al study were classified as experiencing sexual dysfunction and responsiveness to sex was a worry to 64% (55). Kalantaridou et al (2008) assessed sexual function score in 143 women with POF following 3 months of physiological oestrogen replacement (100mcg oestradiol patch) (59). Despite adequate oestrogen replacement 7% of the POF group, compared with 2% of the control group, had an abnormal composite sexual function score ($p<0.001$). The authors do not report the scores at the start of the hormone replacement period, although prior to starting HRT there was only a very short 2 week washout period from any previous regimens used so this data would be of limited value. 15% of women with POF had a free testosterone below the normal limit and 5% had total testosterone below normal; both these groups had a trend towards lower sexual function scores. There was a significant correlation between sexual function score and total testosterone in POF ($p=0.03$) and a trend to correlation with free testosterone ($p=0.06$). These associations were not seen in the control group. Van der Stege et al (2008) conducted a questionnaire study of 81 women with POF (response rate 41%) (15). Women with POF were found to be less satisfied with sexual aspects of their life and had less sexual motivation than controls. Women with POF taking oestrogen replacement were found to have lower levels of sexual satisfaction than those not using oestrogen replacement. However, there could be a bias towards women with more significant symptoms opting to take treatment. Total testosterone increased the frequency of desire for sexual contact but no other parameters. Both of these studies measured testosterone levels but the value of this in isolation is uncertain; it is recognised that androgen metabolism is complex and serum levels may not reflect tissue levels (60). However, several studies have found that testosterone levels are reduced in POF (61). Another recent cross-sectional study confirmed high rates of sexual dysfunction in POF: 62% in the POF group (n=58) and 38% in the control group (n=58) were classified as having sexual dysfunction ($p=0.009$) (62).

2.2.2 Long term complications of POF

2.2.2.1 Osteoporosis

2.2.2.1.1 Definition and DXA

The World Health Organisation (WHO) defines osteoporosis as 'a systemic skeletal disease characterised by low bone density and micro-architectural deterioration of bone tissue with consequent increase in bone fragility' (63). Diagnosis is made by a T score ≤-2.5 measured by dual energy X-ray absorptiometry (DXA) scan at the lumbar spine, hip or radius (forearm). The T score is the number of standard deviations the bone mineral density (BMD) differs from the young adult mean, adjusted for gender and ethnicity. Central DXA (lumbar spine or hip) is preferred as it assesses the sites of most clinical relevance and is proven to predict fracture risk (64). The most powerful predictor of fracture is hip DXA. The best site for monitoring response to treatment is the lumbar spine due to good precision and relatively large changes with treatment (64).

	T-score
Normal	T ≥ -1.0
Osteopaenia	-2.5 < T < -1.0
Osteoporosis	T ≤ -2.5
Established osteoporosis	T ≤ -2.5 plus one or more fragility fractures

Table 2 World Health Organisation definitions of osteopenia and osteoporosis

The Z score is the number of standard deviations the BMD differs from an age, gender and ethnicity matched mean. It therefore gives a measure of the patient's BMD compared with that expected for his or her age group. Z scores are used in preference to T scores in individuals who have not yet reached peak bone mass.

2.2.2.1.2 Clinical consequences of osteoporosis

The clinical consequence of osteoporosis is fracture, which occurs at lower trauma in osteoporotic bone than in healthy bone. For each standard deviation the BMD drops below the young adult mean, the risk of sustaining a fracture doubles (65). The most common fractures associated with osteoporosis are those of the hip, vertebrae and wrist (63). All are associated with a reduction in quality of life, with a higher impact caused by hip fracture, vertebral fracture or fracture at more than one site (66, 67). Excess mortality rates in the year following a hip fracture, compared with the general population, have been reported as up to 36%. Most studies report at least a doubling of mortality risk which persists for several years after the fracture (68).

BMD is not the only predictor of fracture; half of all fractures occur in individuals who do not have osteoporosis (69). Fracture risk appears to depend on bone quality as well as density and many risk factors have been identified. An assessment of absolute risk of fracture can be estimated in women over 40 using the FRAX™ model, which has been developed from meta-analysis of risk factors and calibrated to the UK population (70). It estimates an individual's risk of hip fracture and major osteoporotic fracture over the next 10 years, based on the following clinical information, with or without femoral neck BMD:

- Age
- Gender
- Body mass index (BMI)
- Prior fragility fracture
- Parental hip fracture
- Current smoking
- Long-term glucocorticoid use
- Rheumatoid arthritis
- Secondary cause of osteoporosis (only included in calculation if BMD unknown)
- Daily alcohol intake 3 or more units

Unfortunately, due to a lack of epidemiological data in younger age groups, the FRAX™ model cannot be used to estimate fracture risk in a person under the age of 40 years.

2.2.2.1.3 Pathophysiology of osteoporosis

Bone mass increases throughout childhood and adolescence and reaches a peak at the age of 20-30. In women the rate of gain in BMD reduces significantly after menarche. The development of osteoporosis depends on both the peak bone mass attained, which is highly variable (63), and subsequent loss.

Determinants of peak bone mass include (71):

- Genetic factors, the exact nature of which are unknown, but are likely to be polygenic
- Gender
- Diet (protein, calcium, vitamin D)

- Endocrine factors, for example GH deficiency, oestrogen deficiency due to anorexia or weight related amenorrhoea, hyperthyroidism

- Physical activity

In the healthy individual, peak bone mass is maintained until the fifth decade when a small amount of bone loss starts to occur. Substantial loss occurs from the time of menopause in women, due to oestrogen deficiency. After 6-10 years the rate of loss slows down (38).

Bone is a dynamic tissue and undergoes remodelling in 3-6 month cycles. Each year approximately 5-10% of the adult skeleton is remodelled. This process repairs microdamage throughout the skeleton and maintains bone structure and strength. In the normal state, bone formation matches resorption. It is controlled by local and systemic hormones and immune cells. Remodelling takes place in bone remodelling units which consist of osteoclasts and osteoblasts. The bone remodelling unit is encased by a 'canopy of cells', which may contribute towards creating the correct microenvironment (72). Osteoclasts resorb bone then osteoblasts lay down organic bone matrix and mineralise it. This is a highly co-ordinated process and the exact control mechanisms are incompletely understood. Osteoblasts have a role in controlling the differentiation of osteoclasts through the expression of osteoclastogenic factors. Macrophage colony stimulating factor (M-CSF) and receptor activator nuclear factor-κB ligand (RANKL) are essential osteoclastogenic cytokines produced by bone marrow stromal cells and osteoblasts. Osteoblasts and bone marrow stromal cells also produce osteoprogenin (OPG) which binds to RANKL receptors and inhibits the effects of RANKL. Immunological factors also play a role: T cells produce interferon, tumour necrosis factor α (TNFα) and RANKL which increase osteoclastogenesis (73). B cells produce OPG. B cell deficiency in mice causes osteoporosis (72). Megakaryocytes in the bone marrow express RANKL and OPG and may also influence remodelling (72). Osteocytes are differentiated osteoblasts which are engulfed in bone matrix. They have long dendrites which extend through the matrix and are believed to play a role in detection of mechanical strain and direction of remodelling (72).

Systemic hormones which act on bone remodelling units include parathyroid hormone (PTH), oestrogen and thyroxine. PTH is secreted by the parathyroid glands in response to a low serum calcium. It acts via osteoblast receptors to induce osteoclast differentiation and therefore increase bone resorption. Hyperthyroidism increases the rate of bone turnover and excess glucocorticoid treatment reduces bone formation (63). Oestrogen has both direct skeletal and extra-skeletal effects, which are discussed below.

2.2.2.1.4 Effect of oestrogen on bone and bone mineral density

Oestrogen acts on bone cells directly and via cytokines. The current understanding of its effects are: (74)

1. Inhibits activation of bone remodelling

2. Promotes apoptosis of osteoclasts and inhibits differentiation

3. Prevents apoptosis of osteoblasts and promotes differentiation

Evidence for inhibition of activation of bone remodelling units comes from histological and bone turnover marker data following administration of oestrogens to postmenopausal women (75). Oestrogen stimulates osteoblastic OPG production which inhibits RANKL and therefore inhibits osteoclastogenesis. Oestrogen deficiency causes an increase in RANKL production by bone marrow stromal cells and osteoblasts, which increases osteoclastogenesis (73). The importance of oestrogen in maintaining bone formation is seen following a short period of oestrogen deficiency, which causes a rapid decrease in bone formation markers (76).

Extra-skeletal effects of oestrogen deficiency are: decreased intestinal calcium absorption, increased renal excretion and an increase in PTH levels (73).

Oestrogen also plays a vital role in the development and maintenence of the male skeleton. Men who are oestrogen deficient from childhood have delayed epiphyseal fusion, low bone density and high bone turnover markers (74). In adult men oestrogen plays a role in regulation of bone formation and resorption and determines bone mass and bone loss (74).

An accelerated phase of bone loss due to oestrogen deficiency occurs in the early post-menopausal years, with subsequent slowing to a rate seen in older men. Numerous studies have shown that HRT prevents bone loss (77). The Women's Health Initiative trial found that HRT reduces hip and vertebral fractures by one third; other fractures were also reduced (78).

The effects of the COCP at the menopause have been much less widely studied. A few small studies indicate that it maintains or slightly increases BMD (79, 80).

2.2.2.1.5 Bone mineral density in premature ovarian failure

Several cross-sectional studies have shown that women with POF have reduced bone density (7, 8, 81). The largest included 442 women aged 18-42 with spontaneous POF (6). 70 concurrent regularly menstruating controls were used along with 353 matched controls from NHANES III (National Health and Nutrition Examination Survey). The POF group were found to have significantly lower BMD at the lumbar spine and hip of 2-3%. 8% of the women with POF had a T score of -2.5 or less (within oesteoporotic range) and 15% had a Z score of under -2 (below the expected range for their age). A Z score of under -2 was significantly associated with: not taking HRT, delay in diagnosis of over a year, onset of irregular periods before age 20, lack of regular exercise, low calcium intake and low serum vitamin D. The authors also looked at the type of prior oestrogen replacement (HRT (n=184), COCP (n=68), HRT/COCP in either order (n=61)) although not the preparation, dose or length of time taken and found no differences in BMD between the groups.

A cross-sectional study which included 4724 postmenopausal women, 582 (21%) of whom had an early menopause (before 45), showed that women with an early menopause have a lifelong

increased risk of fracture (odds ratio for fracture 1.5; CI 1.2-1.9) (10). A recently reported retrospective observational study followed 390 women from the ages of 48 to 84 and assessed bone density, fractures and mortality from all causes (82). The authors found that those with menopause before the age of 47 had significantly higher risks of osteoporosis at the age of 77 (RR 1.83; 95% CI 1.22-2.74), risk of fragility fracture by the age of 84 (RR 1.68, 95% CI 1.05-2.57) and mortality by 84 (RR 1.59; 95% CI 1.04-2.36) compared with those who had menopause later than 47 years.

To date there is only one published prospective trial which has investigated bone density with different oestrogen preparations in women with POF (83). This was a small open-label randomised crossover trial. Following a 2 month 'washout' period participants were randomised to either 'physiological' or 'standard' hormone replacement. The physiological regime consisted of transdermal oestradiol (100mcg week 1 and 150mcg weeks 2-4) with 200mg progesterone vaginally twice a day in weeks 3-4. The standard regime was equivalent to a standard strength COCP – 30mcg ethinyloestradiol with 1.5mg norethisterone daily for weeks 1-3 then a pill free week in week 4. The authors report a significant increase in Z score following a year's treatment with the physiological regimen but no difference with the COCP. However, when the change in BMD was directly compared between the two groups no significant difference was found. Interestingly, although both regimens suppressed the marker of bone resorption (CTX), the markers of bone formation (bone ALP and P1NP) were increased following the physiological treatment but decreased following COCP. CTX was suppressed significantly more by the COCP than by the physiological regimen. Both treatments suppressed FSH and LH by a similar amount. The drop-out rate in this study was high with only 18/34 completing the trial. Most of the women did not have spontaneous POF – of those who completed 4 had POF secondary to cancer treatment, 7 had Turner syndrome and 7 had idiopathic/surgical POF (the exact number with idiopathic POF is not provided).

Bone health is reported as a major concern to women with POF; 92% in a recent questionnaire study were concerned about it, which was as high as the number with fertility concerns (55).

2.2.2.1.6 Markers of bone turnover

Bone turnover markers are divided into markers of bone formation and markers of bone resorption. They change more quickly than BMD following initiation of treatment and are indicative of future BMD changes (84). Bone resorption markers are independent predictors of fracture risk (85).

P1NP (procollagen type I N-terminal propeptide) and CTX (cross-linked C-terminal telopeptide) are markers of bone formation and resorption respectively. More than 90% of the organic bone matrix is made up of type 1 collagen. This is constructed from type 1 procollagen and P1NP is a specific marker for type 1 collagen deposition, increasing when new bone is formed. CTX is a type 1 collagen fragment and so levels are increased when bone resorption increases. The other frequently used bone formation marker is bone alkaline phosphatase (bone ALP).

At the menopause, bone turnover markers increase and remain elevated. A reduction of up to 70% is seen with HRT (85). The effect of the COCP has been less widely studied but it has been reported to decrease markers of bone turnover in women aged 35-49 (86). A very short (6 month) crossover study of 17 women with Turner's syndrome found the COCP to decrease bone formation markers whereas HRT did not (87). The crossover study described above found that both the COCP and HRT decreased CTX but markers of bone formation were increased by HRT and decreased by the COCP (83).

2.2.2.2 Cardiovascular disease

For many years, it was considered that HRT decreased the risk of developing cardiovascular disease (CVD), based on data from observational studies, including the large Nurses' Health Study involving 116,258 women, in which a 33% reduction was seen in those initiating HRT close to the menopause (88). This association was not confirmed in the initial Women's Health Initiative (WHI) publications, which reported that combined HRT increases coronary heart disease by 29% (78). The WHI was a large randomised controlled trial investigating HRT as primary prevention for CVD. The mean age of the trial participants was 63 years, 34% had a BMI of over 30 and HRT was commenced without a clinical reason, at a higher dose than would usually be used in this older age group. Therefore, the applicability of these results to the general early postmenopausal population has been widely questioned, and the results cannot be generalised to women with POF.

Oestrogen is considered to confer no overall benefit to women with established cardiovascular disease and may cause an early increase in risk. This was evaluated in the Heart and Estrogen/Progestin Replacement Study (HERS) which randomised 2763 women with cardiovascular disease to either continuous combined HRT or placebo (89). After a mean follow up of 4.1 years there was no difference in myocardial infarction or death secondary to coronary heart disease. Although there was no overall difference in events between the groups, in the HRT group there were more events in years 1 and 2 and fewer in subsequent years, suggesting an early increase in risk with possible later benefit.

The WISDOM trial (Women's International Study of long Duration Oestrogen after Menopause) (90) initially aimed to investigate the use of HRT in younger women (age 45-60), which is the usual age of commencing HRT in clinical practice, for 10 years. However, this was extended to 69 years and the final mean age of participants was 63 years. Findings were similar to the WHI trial, with an increase in cardiovascular and thrombo-embolic events in the combined HRT compared with the placebo group. The trial was stopped early following the initial publication of the WHI results in 2002. 4385 women were randomised (26% of the original target) and followed for a mean of 11.9 months. The short follow up and advanced age of participants meant that trial was unable to answer the original question of quantifying the health risks and benefits for women who start HRT around the time of the menopause. A trial large enough to evaluate this specifically has not yet taken place and is not planned. However, a meta-analysis of 23 randomised controlled trials of HRT (total 39,049 participants) demonstrated a significant

reduction in coronary heart disease in women starting HRT before the age of 60 or within 10 years of the menopause (91). The same team have more recently evaluated the relationship between hormone therapy and mortality in women under 60 using data from 19 randomised trials involving 16,000 women and found a significant reduction in mortality in women using HRT (92).

A re-analysis of the WHI data showed that in women under 60 and those who commenced HRT within ten years of the menopause, there was a trend towards cardiovascular benefit (93). The trial was discontinued before statistical significance was reached. This correlates with results from the meta-analysis described above and with knowledge about the direct effects of oestrogen on the cardiovascular system. Results from the Danish Osteoporosis Prevention study also support this. This was an open-label trial in which women with an average age of 50 and time since menopause of seven months were randomised to HRT or no treatment. After ten years, there was a reduction in cardiovascular events in the HRT group (hazard ratio 0.48; 95% CI 0.26 to 0.87) (94). Oestrogen has been reported to have a beneficial effect on endothelial function in younger post-menopausal women, and to reduce atherosclerosis in oophorectomised monkeys (95). There has been one cohort study which aimed to assess the association between early menopause, taking oestrogen replacement and the risk of ischaemic heart disease (96). This study included 10533 postmenopausal Danish nurses aged over 44 in 1993 (86% response rate) who gave their age at menopause via questionnaire and were then followed up via the national registry for cardiovascular events. The findings were that menopause both at under 40 and under 45 were associated with increased ischaemic heart disease; this was most pronounced with surgical menopause but also seen in POF. In women with surgical menopause who took HRT there was no increased risk of IHD compared with women who did not have early menopause.

2.2.2.2.1 Cardiovascular disease in premature ovarian failure

Epidemiological studies have associated an earlier menopause with increased cardiovascular risk. A recent meta-analysis reported that the risk of developing CVD is increased by 38% in women who experience an earlier menopause compared with those who reach it at the age of 50 (11). A previous study suggested that for each year the menopause is delayed, the risk of cardiovascular mortality decreases by 2% (12). No studies to date have examined the effect of an early menopause and oestrogen replacement on direct CVD outcomes. It is recommended that women with POF take hormone replacement to reduce the risk of CVD (97) but this is not evidence-based.

Kalantaridou et al (2004) investigated endothelial function in 18 women with POF (98). The POF group had significantly lower flow-mediated vasodilation (FMD; a measure of the function of the endothelium in response to arterial occlusion) than controls. This was normalised after 6 months of HRT. The COCP was not investigated. Another study investigating FMD in 20 women with POF not taking any hormone replacement also found that FMD was reduced compared to age-matched controls; no HRT was given in this study. The POF group in this

study also had a lower level of circulating endothelial progenitor cells and increased carotid intima media thickness compared with controls, as well as impaired left ventricular diastolic function assessed by echocardiogram (99). Goldmeier et al compared the endothelial function of 17 women with POF taking daily conjugated oestrogens and cyclical medroxyprogesterone acetate with 15 controls. FMD was similar in the two groups but the POF group had slightly higher diastolic and systolic mean arterial pressures, although only the diastolic pressure was significantly different. The POF group also had impaired baroreflex sensitivity and reduced heart rate variability (measures of autonomic response and possible markers of cardiovascular disease) compared with the control group (100).

Langrish et al (2009) conducted a crossover trial to assess the effects of two different hormone replacement regimes in POF on blood pressure and renal function (101). Loestrin 30 (a COCP containing ethinyloestradiol and norethisterone) was compared with an oestradiol patch plus cyclical progesterone in a crossover trial. Blood pressure was found to be slightly lower with the oestradiol patch compared with COCP. However, follow up was poor with 18/42 completing the study. Very few of the participants had spontaneous POF; most participants had chemotherapy-induced POF or Turner syndrome. The type of progesterone given with the oestradiol was also variable, depending on participant choice of route of administration.

2.2.2.2.2 Lipids

After the menopause, there is an unfavourable change in lipid profile, with an increase in total and LDL cholesterol and triglycerides (TG) and a decrease in HDL cholesterol (102).

It has been proposed that a change in lipid profile contributes to the increased cardiovascular risk associated with POF. A cross-sectional study of 90 women with POF found slightly higher TG and borderline lower HDL compared with controls. No relation to serum oestradiol levels or duration of oestrogen deficiency was found (13). The effect of 6 months' oestrogen treatment on lipid profile in POF has been investigated in two small studies. A crossover study of 17 women with Turner's syndrome comparing the effects of an ethinyloestradiol-containing COCP with an HRT preparation of conjugated equine oestrogen and medroxyprogesterone acetate found no difference in lipid profile following 6 months of either treatment (87). In their study of 18 women with spontaneous POF, Kalantaridou et al (2004) found that there was no difference in the baseline lipid profile compared with controls, but after 6 months of HRT the LDL significantly decreased and TG increased compared with pre-treatment values although there was still no significant difference from the control group (98).

2.2.2.2.3 C-reactive protein (CRP)

CRP is an acute inflammatory marker. Part of the pathogenesis of atherosclerosis and subsequent cardiovascular events is considered to be inflammatory and therefore CRP has been used to attempt to quantify cardiovascular risk. High sensitivity CRP (hsCRP) is a strong predictor of cardiovascular events, even after adjusting for established risk factors, and has been suggested by some to be the best marker for future CVD in women (103, 104). CRP

increases with hormone replacement therapy but remains a predictor for CVD (105). However, other studies have questioned the value of CRP as an independent cardiovascular risk factor. One study investigating cardiovascular risk factors with oestrogen treatment found that CRP was increased whereas other markers such as interleukin-6 are not, leading to the question of whether the raised CRP is due to hepatic metabolism of the oestrogen rather than endothelial dysfunction (106). This concern has been raised previously (107). A recent study investigated baseline CRP levels and progression of atherosclerosis in 423 postmenopausal women over almost 3 years (108). There was no relationship between initial CRP levels and progression of atherosclerosis, although a relationship between CRP and cardiovascular death/myocardial infarction was demonstrated. Interestingly, CRP has also been reported as variable through the menstrual cycle (109).

2.2.3 Fertility

Loss of fertility is reported as one of the most distressing aspects of POF, especially with increasing age (14, 55, 110). However, spontaneous pregnancy following diagnosis is estimated to occur in 5-10% (3). There are many published case reports on pregnancy in POF. These include reports on the occurrence of pregnancy with no treatment (n=11), oestrogens (n=32), clomiphene citrate (n=1), corticosteroids with oestrogens (n=3), gonadotrophins (n=1) and gonadotrophin releasing hormone agonists with human menopausal gonadotrophin (n=2) (3). A large number of reported pregnancies were in women not taking any treatment, leading to doubt over whether the interventions actually played a role in those women receiving treatment who became pregnant.

There has been little good quality research in this area. Interventions assessed in controlled trials include oestrogens with gonadotrophin releasing hormone agonists or danazol; gonadotrophin releasing hormone agonist plus gonadotrophins and oestrogen replacement therapy. None produced a higher pregnancy rate than the background observed rate. Studies to date are limited by their small size, short follow up periods, wide variations in inclusion criteria and few randomised controlled trials. Many studies have used ovulation rather than pregnancy as an endpoint.

It is impossible to accurately predict which women will conceive spontaneously. A shorter duration of amenorrhoea has been identified as a positive prognostic factor (3). Autoimmune aetiology of POF and the presence of ovarian activity on ultrasound scan have also been suggested to be associated with a higher chance of spontaneous conception (111).

Dehydroepiandrosterone (DHEA) has recently been postulated as a treatment for POF and has been claimed in individual cases to reduce FSH levels and lead to pregnancy (112, 113). It is an androgen synthesised in the adrenal gland and ovary and is a precursor for ovarian sex hormone synthesis. However, there have been no controlled studies to assess its efficacy in POF, although one is underway (114), and there is no theory on its potential mode of action. It is a prescription-only medication in the UK and is not widely available. In America it can be bought over the counter as a food supplement.

The only effective fertility treatment option is in vitro fertilisation with egg donation. This carries good outcomes, dependent on the age and appropriate selection of the egg donor. Clinical pregnancy rates following egg donation are reported to be around 30% (115). A small study in which one group consisted solely of women with POF (36 women) demonstrated an average requirement of 1.75 cycles for a live birth and a cumulative pregnancy rate of 75% (116).

Fertility preservation for women at risk of developing POF is a topical area. Currently, embryo freezing carries the highest chance of success but this is only an option for women with a partner or who wish to use donated sperm. Oocyte freezing has low rates of success of 5-10% chance of a live birth (117) but this may improve with newly developed techniques of rapid freezing. A case of a 14 year old with mosaic Turner syndrome opting for oocyte freezing is described (36). Ovarian tissue cryopreservation prior to chemotherapy or radiotherapy and auto-transplantation following treatment has recently been developed and pregnancies are reported (118).

As described above, follicular activity and ovulation are observed in POF (3, 4), making it different from the 'normal' menopause. The term 'ovarian reserve' indicates the number and quality of remaining oocytes. The only way to definitively determine this would be to examine the whole ovary histologically. Obviously this is not desirable and so indirect markers of the ovarian reserve have been developed. These markers are used in assisted conception to predict response to ovarian stimulation. Potential uses in POF include diagnosis, predicting women who may respond to fertility treatment, predicting future ovarian function and improving understanding of pathogenesis. To date, few studies have investigated a combination of markers of ovarian reserve in POF and none have examined any markers longitudinally.

2.2.3.1 Ovarian reserve markers

The traditional marker of ovarian reserve is serum follicle stimulating hormone (FSH) concentration which is produced by the pituitary and stimulates the growth of new follicles in the ovary. However, FSH is not an ideal marker: it must be correctly timed in order to be interpreted, it has a high cycle-to-cycle variability, only increases in the later stages of ovarian ageing and is poor at predicting reproductive status (119). In one study, 11% of 112 women with confirmed POF and not currently taking oestrogen replacement had an FSH level lower than the criteria for diagnosis (120). Attention has turned to more direct markers, originating from the ovary itself.

2.2.3.1.1 Antral follicle count

In a normal monthly menstrual cycle, approximately ten oocytes from each ovary are selected to develop as follicles. These are called antral follicles and can be counted on trans-vaginal ultrasound scan. One follicle is eventually selected to become the dominant follicle. The dominant follicle continues to grow and will ovulate in response to a surge of luteinising hormone (LH) when it has reached an appropriate size. The antral follicle count (AFC) is considered to be related to the total number of oocytes, although this has not been definitively

proven. It is a strong predictor of ovarian response to stimulation and the pattern of decline in AFC with age fits with the decline in numbers of primordial follicles (121). The AFC is independent of the time of the menstrual cycle (122, 123).

2.2.3.1.2 Ovarian volume

Ovarian volume decreases with age and has therefore been used as a marker of ovarian reserve (124). Compared with AFC, it is a poorer predictor of ovarian response to stimulation (125). However, it is easy to measure and may be useful in contributing to predictive models in POF (126).

2.2.3.1.3 Anti-Müllerian hormone

Anti-Müllerian hormone (AMH) is a glycoprotein peptide growth factor and a member of the TGF-β (transforming growth factor-beta) family. Its name is derived from its role in the male fetus, where it is produced by Sertoli cells to cause regression of the female Müllerian duct system. AMH is expressed in antral follicles from the stage of primordial follicle recruitment until selection for dominance. It is therefore indicative of the number of growing antral follicles (127). Mouse studies suggest that plasma AMH acts as an intra-ovarian signal for the size of the antral follicle pool and enables its regulation; lack of AMH causes accelerated follicular recruitment and exhaustion (128). Unsurprisingly, AMH is related to AFC. AMH can be measured at any stage of the menstrual cycle and unlike AFC is completely operator-independent. AMH is as good as AFC at predicting poor response to ovarian stimulation in IVF (129).

There is interest in using ovarian reserve markers, and in particular AMH, to predict the age at menopause. The theory is that, in a similar way to the menopause occurring when the threshold of around 1000 follicles is reached, there is a threshold AMH level at which menopause occurs. If the pattern of decline of AMH is known then the age at menopause could theoretically be predicted. A cross-sectional study investigating serum AMH in 144 fertile women demonstrated a relationship between AMH levels and the population's pattern of age at menopause (130). However, to date no ovarian reserve marker has been found to accurately predict age at menopause (131). The study included few young women making predictions of menopausal age based on AMH for young women even more uncertain.

2.2.3.1.4 Inhibin B

Inhibin B is also a glycoprotein of the TGF-β family. It is produced by granulosae and thecal cells in small antral follicles and inhibits FSH secretion. Its concentration is related to the mass of granulosa cells, leading to its use as a marker of ovarian function. Inhibin B increases until the mid-follicular phase and then decreases in inverse correlation with FSH; measurement must therefore be correctly timed at the beginning of the follicular phase. Inhibin B serum concentrations decrease before FSH increases with age-related decline of ovarian function. Inhibin B appears to be a poorer marker of ovarian reserve than AMH and AFC (132).

2.2.3.2 Ovarian reserve markers in POF

Few studies have been carried out in this area. An early study (in 2006) measured serum AMH levels and small antral follicles in 12 women with POF and found both to be significantly reduced compared with healthy controls and women with hypogonadotrophic hypogonadism (133). 10 women with POF had undetectable AMH levels and 2 had very low levels (133). Knauff et al (2009) took a single measurement of AMH, inhibin B and AFC in 112 women with POF (FSH over 40IU/l), as well as groups with a regular cycle and with oligomenorrhoea, both with raised FSH (over 10.2 IU/l) (120). They reported that AMH was a superior marker of the clinical stage of ovarian failure compared with AFC; inhibin B was the poorest. AMH was detectable in 5% of women with POF and was below the 5th centile in all of these. In the group with an FSH of over 10.2IU/l and oligomenorrhoea, 33% had a normal AMH; in the group with FSH of over 10.2IU/l and regular cycles, 75% had a normal AMH. This leads to the theory that in some women who are developing spontaneous POF there is still a significant remaining ovarian reserve. Another study which evaluated the AMH levels of 147 women with POF found a normal level in 5%, a level below normal in 18% and undetectable levels in 77% (24).

Massin et al (2008) postulated that there may be either an absence of follicles or impaired follicle growth in POF (134). They compared ovarian histology with serum inhibin B and AMH and found both to be predictive of the presence of follicles. However, there is a flaw in regarding ovarian histology as a 'gold standard' for the presence of follicles: pregnancies in the absence of follicles on histology are well documented.

No research to date has looked longitudinally at the progression of ovarian reserve markers in POF. In order to rely on these markers for diagnosis, or to develop ways of using them in a fertility setting, we need to know this. This study begins to address this, and in looking at several markers will aim to examine their inter-relation in this population.

2.2.4 Treatment of POF

Oestrogen replacement is recommended in POF to protect against bone loss and cardiovascular disease and alleviate symptoms. The most suitable form is unknown and management is not currently evidence-based (1, 16, 17). Currently, either HRT or the COCP is commonly prescribed. A recent survey of 130 members of the British Menopause Society revealed that 56% prescribe HRT in preference to COCP and 23% have no preference (135). HRT is considered to be more 'physiological' but some young women dislike taking a preparation associated with older women. The COCP contains synthetic oestrogen at a higher dose with a 'pill free' week. It may be considered more 'peer friendly' than HRT (it is taken by 16% of women aged 16-49 (136)) and there is no prescription charge. The quantification of the relative advantages and disadvantages of each treatment in POF is important in order to provide evidence-based advice on treatment options. Affected women face many years of oestrogen replacement and deserve more information than is currently available.

There are also a significant proportion of women who decline treatment. To date, there have been no prospective studies in POF investigating the quantitative effects on bone density or other outcomes of different treatments compared with no treatment. In spite of the almost universal recommendation of oestrogen replacement, the number of women who decline treatment may be higher than often perceived. There are no published studies estimating the number of women who decline oestrogen treatment in the long term. A study in America investigating bone density in 442 women with POF included 92 (21%) who had never taken oestrogen replacement (6). Another American study investigating 50 well-educated women with Turner's syndrome found that only 68% were taking oestrogen replacement therapy in accordance with current guidelines (137). Of the 16 women not taking oestrogen replacement, 11 were not compliant with treatment and 5 had not been prescribed oestrogen replacement. The most important factor in predicting compliance with treatment was education about oestrogen replacement and bone health. Another study of 89 women with POF due to chemotherapy or radiotherapy, of whom 43 were taking oestrogen replacement, asked an open-ended question regarding views on oestrogen therapy (56). 26% were concerned about the risk of cancer, especially breast cancer; this may be high due to the nature of the group. Concerns over weight gain were cited by 13%. Reasons for taking oestrogen therapy included protection against osteoporosis (48%), replacement of hormones (29%) and relief of menopausal symptoms (24%). It has been reported that many women with POF believe that adverse media reports regarding HRT apply to their age group (138). Even after the provision of accurate advice, this may contribute towards a decision to decline treatment.

2.3 Microgynon 30 and Nuvelle

Microgynon 30 is a COCP containing ethinyloestradiol 30mcg and levonorgestrel 150mcg. One pill is taken daily for 21 days followed by a 7 day break which induces a withdrawal bleed. It is a popular first choice COCP. Ethinyloestradiol is a synthetic oestrogen contained in most COCPs. Levonorgestrel is a synthetic 19-nortestosterone derived progestogen.

Nuvelle is a cyclical HRT consisting of oestradiol 2mg daily, with the addition of levonorgestrel 75mcg for 12 days a month, producing a monthly bleed. 'Natural' oestrogens such as oestradiol are increasingly popular. Nuvelle was chosen for this study because it also contains levonorgestrel.

Chapter 3 Methods

3.1 Regulatory approvals

Prior to starting the trial, I obtained Ethics, Medicines Healthcare Regulatory Agency (MHRA) and local Research and Development (R&D) approvals. I completed the necessary forms via the Integrated Research Application System (IRAS) and EudraCT websites and attended Guy's Research Ethics Committee meeting to answer queries. Two small adjustments were recommended by the committee prior to final Ethics Committee approval.

3.2 Recruitment

Participants were recruited through the Premature Ovarian Failure (POF) and Menopause Clinics, the Reproductive Medicine Clinic, local GPs and The Daisy Network (a national patient-run support group for women with POF).

All potentially eligible women who had previously attended the POF/Menopause clinics were invited to participate by letter and telephone. Women who attended as new referrals were approached in the clinic. I wrote to local GPs to encourage them to refer any women who may be interested in participating.

I visited local hospitals to publicise the trial. The Daisy Network advertised the trial on their website and I attended their 2009 annual conference to talk about the trial. The trial advertisement was also posted on the 'Menopause Exchange' website and published in their newsletter. A circular e-mail was sent to King's College London employees. In November 2010, the Joint Clinical Trials Office issued a press release to aid recruitment. This resulted in coverage on BBC online, Women's Weekly and the Daily Express. The trial advertisement also featured on the Guy's and St Thomas' Foundation Trust Hospital intranet.

Women interested in participating were provided with written information (the Participant Information Sheet) and given time to consider participation. If appropriate and when possible, reasons for non-participation were sought and recorded. Visits were arranged in the morning so that questionnaires were completed at a similar time at each visit and participants could fast prior to blood samples.

3.3 Protocol

Following written informed consent, women were seen for a screening visit. A full medical history was taken and physical examination performed. Blood was taken for repeat follicle stimulating hormone, thyroid function tests and fasting lipid profile as necessary. Once the inclusion/exclusion criteria (see appendix 1) had been confirmed, the participant was enrolled into the study and assigned a study number.

Each participant was seen at baseline and at 3, 6, 12, 18 and 24 months. At the baseline visit, the participant decided whether she wanted to be in the 'no treatment' or 'active treatment' group. Participants in the 'active treatment' group were randomised to Microgynon 30 (ethinyloestradiol 30mcg and levonorgestrel 150mcg taken for 21 days out of 28) or Nuvelle

(oestradiol 2mg daily, plus levonorgestrel 75mcg for 12 days a month) using the secure internet randomisation website www.sealedenvelope.com. It was not feasible to blind the participant or the candidate to the medication and this was therefore an open label study. A supply of the study medication in its original packaging was provided at 6 monthly intervals. The visit schedule below shows the investigations performed at each visit.

All visits took place in the morning to allow an overnight fast prior to the blood test and to enable the questionnaires to be done at the same time of day on each occasion. Unfortunately it was not possible to see participants at the same stage of the cycle for each visit, due to the constraints of organising bone scans and the flexibility needed by the participants, most of whom were working full time or looking after young children. At each visit, enquiries were made about adverse events and changes to medications or new medications, including over the counter preparations. General multivitamins were allowed but women were asked not to take any herbal preparations or high-dose calcium tablets during the course of the trial.

Summary of visit schedule

Visit	Pre-trial	0 (start)	3 months	6 months	12 months	18 months	24 months
Medical history	x						
Medical examination	x						
Obtain written informed consent	x						
Review in clinic	x	x	x	x	x	x	x
Height	x						
Weight	x	x	x	x	x	x	x
Decide treatment or no treatment		x					
Blood pressure	x	x	x	x	x	x	x
Randomisation – HRT or COCP		x					
Dispense and collect medication		x		x	x	x	
Questionnaires (MS-TSQ[1], B-PFSF[2] MGCS[3], SF 36[4] and PHQ-9[5])		x	x	x	x	x	x
DXA bone scan		x[6]		x	x		x
Trans-vaginal ultrasound scan		x		x	x		x
Blood test	x[8]	x [7]		x [7]	x [7]		x [7]
Adverse event monitoring		x	x	x	x	x	x

[1]Menopause Symptoms–Treatment Satisfaction Questionnaire
[2]Brief Profile of Female Sexual Function
[3] Modified Greene Climacteric Scale
[4]Short Form-36 version 2
[5]Patient Health Questionnaire–9
[6]If a dual energy X-ray absorptiometry scan was performed within the last 3 months, it was not repeated
[7]lipid profile/high sensitivity C-reactive protein/ C-terminal cross-linked telopeptide / procollagen type I N-terminal propeptide /inhibin B/anti-Mullerian hormone
[8]follicle stimulating hormone, thyroid stimulating hormone, lipid profile

Table 3 Summary of visit schedule

TREATMENT OF PREMATURE OVARIAN FAILURE TRIAL - flowchart

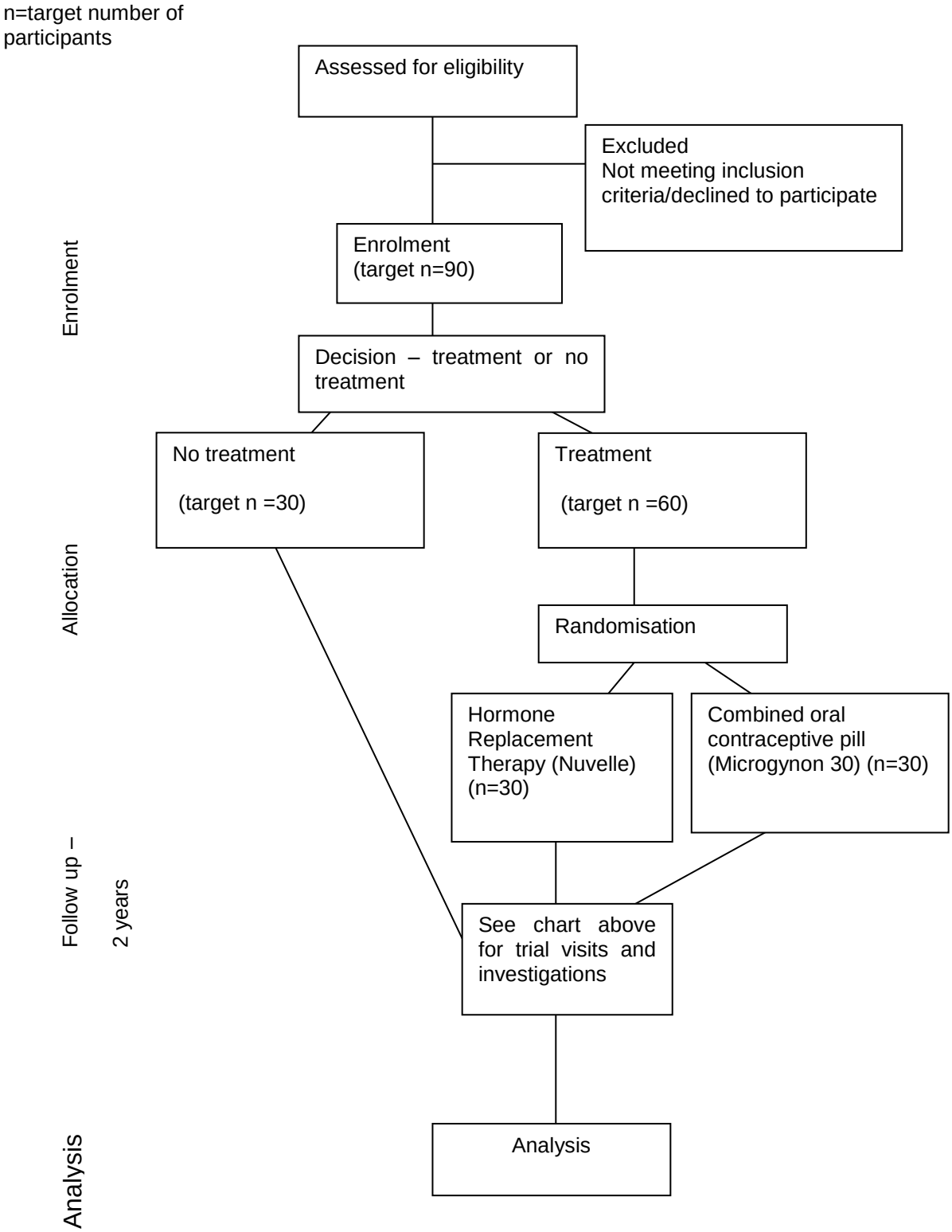


Figure 1 Treatment of Premature Ovarian Failure Trial flowchart

3.4 Dual energy X-ray absorptiometry scans

Central skeleton (hip and lumbar spine) dual energy X-ray absorptiometry (DXA) measurements are considered to be the gold standard for measuring bone mineral density (BMD) and for monitoring changes in BMD following treatment (64). The lumbar spine was chosen as the primary outcome for this study because it is the most sensitive to changes and considered the best site for monitoring changes with treatment (64, 140). We recognise that degenerative changes in the lumbar spine can have an impact on bone density and could potentially progress over two years. However, in this young population with a mean age of 40 we considered the risk of the progression of significant degenerative changes affecting the results to be minimal. Bone markers were not chosen as the primary outcome due to their high variability and lower evidence of relation to fractures than central BMD.

Bone density at the lumbar spine (L1-4) and hip was measured by DXA scan (Hologic Discovery model) in the Osteoporosis Research Unit at Guy's Hospital. The scans were performed by a small number of radiographers, which minimises operator variations. A DXA scan of the lumbar spine and hip takes approximately 10 minutes, including positioning. The very low radiation dose of 8 microsieverts (0.008 mSv) is equivalent to approximately two days' of everyday background radiation.

The principle behind DXA scanning is the transmission of high and low energy photon X-rays through the body (141). The dual energy X-rays are produced by a fan beam DXA system which performs a single sweep across the patient over a few seconds. The amounts of high and low energy X-rays transmitted are recorded and this enables the densities of two different types of tissue (bone and soft tissue) to be calculated. The machines are calibrated each morning using phantom models, with a co-efficient of variation over the course of the study of 3.4% (see appendix F). In vivo reproducibility is good, with a coefficient of variation for lumbar spine and total hip BMD of 1-1.5% (142).

The T-score is calculated from the mean healthy young adult BMD and standard deviation, and the Z score from age, gender and ethnicity-specific BMD and standard deviation. Data for this is derived from the Hologic database (for lumbar spine) and NHANES III data (for femoral neck and total hip). In women under 50, Z rather than T scores should be used to interpret DXA results (143). The Z scores for the local population at Guy's Hospital are slightly higher than those from the US-derived data (+0.33 at the spine and +0.39 at the hip) but the US-derived data is still used to calculate them in order to enable comparisons to be made (144).

3.5 Blood samples

3.5.1 Acquisition

Blood samples were taken from the antecubital fossa using a vacutainer system following an overnight fast of at least 14 hours. One serum bottle was sent to the laboratory at Guy's and St Thomas' NHS Foundation Trust for analysis of high sensitivity C-reactive protein (hsCRP) and lipid profile. Immediate analysis of the lipid profile is more accurate than analysis of previously frozen samples. Another serum sample and one plasma sample were spun immediately in a

chilled centrifuge at 3000 revs/minute for 10 minutes. The samples were then separated and frozen in aliquots of 1ml for storage at -20°C (3 samples of serum for C-terminal cross-linked telopeptide / procollagen type I N-terminal propeptide /Inhibin B and one sample of plasma for anti-Mullerian hormone. Frozen samples were transported on ice to St Thomas' Hospital every three months for storage at -80°C until analysis by GSTS Pathology at the end of the trial.

3.5.2 Analysis

3.5.2.1 Markers of bone turnover

3.5.2.1.1 procollagen type I N-terminal propeptide

Serum procollagen type I N-terminal propeptide (P1NP) was measured by the Roche Elecsys total P1NP test. This is an automated assay using anti-P1NP monoclonal antibodies to detect both trimeric and monomeric P1NP. The sample is first incubated with biotintylated monoclonal P1NP-specific antibody. It is then incubated with streptavadin labelled microparticles and a monoclonal P1NP-specific antibody labelled with a ruthenium complex. These form a sandwich complex. The mixture is put into a measuring cell and the microparticles are magnetically captured onto an electrode. When a voltage is applied the chemiluminescent emission can be measured and plotted onto a calibration curve. Serum P1NP is a stable when stored frozen and is unaffected by food intake (145).

3.5.2.1.2 C-terminal cross-linked telopeptide

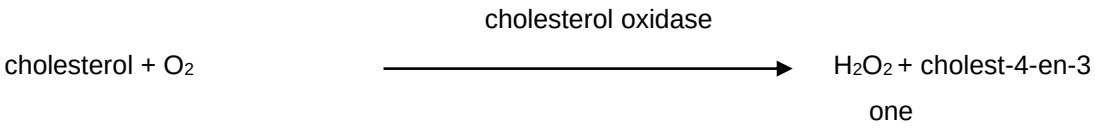
Serum C-terminal cross-linked telopeptide (CTX) was measured by the Roche Elecsys β-CrossLaps/serum assay, which is an automated assay using two specific monoclonal antibodies. The principle of the assay is the same as the P1NP assay. It is also stable when frozen, but needs to be collected fasting, because changes over the period of the day are seen (146).

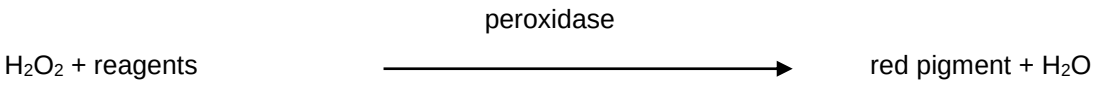
3.5.2.2 Lipid profile

The total cholesterol, high density lipoprotein (HDL) and triglyceride (TG) values were measured using enzymatic colimetric assays on a Roche automated clinical chemical analyser. The principle of each assay is the same. For total and HDL cholesterol measurement, enzymes are added to catalyse the reaction of cholesterol ester plus water to cholesterol plus fatty acid, and then the reaction of cholesterol plus oxygen to hydrogen peroxide. Reagents are then added to produce a dye. The intensity of this dye is directly proportional to the cholesterol's concentration and is measured photometrically. A similar technique was used to quantify triglycerides. The reactions for each are detailed below.

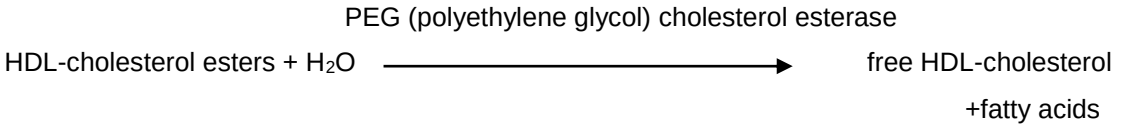
3.5.2.2.1 Total cholesterol

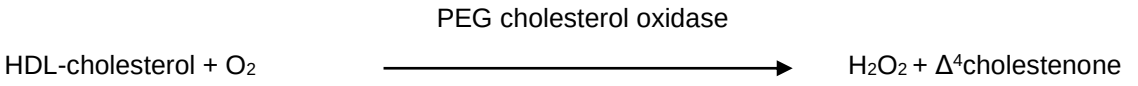
$$\text{cholesterol esters} + H_2O \xrightarrow{\text{cholesterol esterase}} \text{free cholesterol} + \text{fatty acids}$$

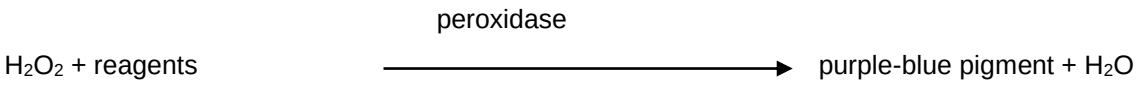




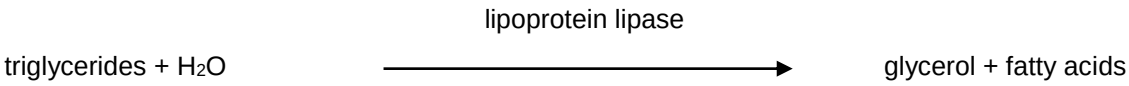
3.5.2.2.2 High density lipoprotein

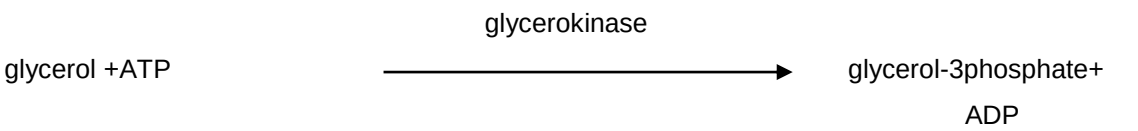


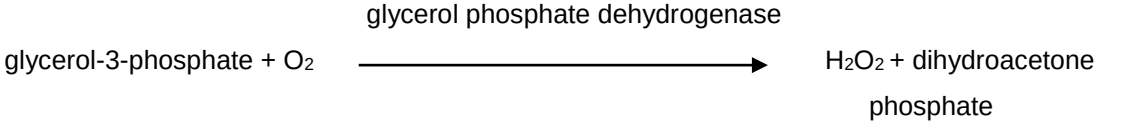


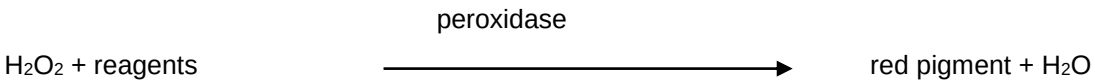


3.5.2.2.3 triglycerides









3.5.2.2.4 LDL

LDL cholesterol was calculated using the Friedewald formula (147):

LDL cholesterol (mmol/l) = total cholesterol – HDL cholesterol – (TG x 0.45)

A baseline lipid profile requiring treatment was an exclusion criterion and therefore the conditions of using this formula were met in all participants.

3.5.2.3 High sensitivity C-reactive protein

High sensitivity C-reactive protein (hsCRP) was measured by particle enhanced immunoephelometry using BM™ Systems. Serum is mixed with polystyrene particles coated with monoclonal antibodies specific to CRP, forming aggregates of polystyrene-CRP. The density of these aggregates is measured by passing a beam of light through the sample; the intensity of the scattered light is proportional to the CRP concentration.

3.5.2.4 Serum markers of ovarian reserve

3.5.2.4.1 Anti-Mullerian hormone

Anti-Mullerian hormone (AMH) was measured using the AMH Gen II ELISA (Beckman Coulter). This is an enzyme-linked immunosorbent assay (ELISA) using a 2-immunological step process. The sample is incubated in a microtitration well coated with AMH antibody. Following incubation and washing, labelled anti-AMH detection antibody is added. After a second incubation and washing, streptavidin-horseradish peroxidase is added and a third incubation and washing takes place. An acidic stopping solution is then added and the degree of enzymatic turnover of the substrate is assessed by dual wavelength absorbance measurement. The sample's AMH concentration is directly proportional to the absorbance. AMH calibrators are used to plot a calibration curve enabling AMH concentrations to be calculated.

This assay replaces the previously used AMH ELISA assays from Diagnostics Systems Laboratory and Immunotech, which gave different values for AMH measurement due to different standardisation. The functional sensitivity of the AMH Gen II ELISA is 1.5 pmol/l (148). AMH readings are approximately 40% higher than those obtained from the DSL ELISA assay. A value of 6.4pmol/l has been identified as predicting poor response to IVF but there are no age-related normative data as yet (148).

3.5.2.4.2 Inhibin B

Inhibin B was measured using the Inhibin B Gen II ELISA (enzyme-linked immunosorbent assay). The principle is the same as that for measuring AMH. The sample is incubated in microtitration wells containing anti-Activin B antibody. After incubation and washing, the wells are incubated with biotinylated anti-Inhibin alpha-subunit detection antibody. Following a second incubation and washing, the wells are incubated with streptavidin labelled with

horseradish peroxidase and the process is as for AMH but using a set of Inhibin B Gen II calibrators to plot the calibration curve so that Inhibin B can be calculated.

3.6 Ultrasound markers of ovarian reserve

The antral follicle count (AFC) and ovarian volume were assessed by trans-vaginal 2-dimensional ultrasound using an 8.0 MHz probe. Ultrasound uses the differential reflection of high frequency sound waves by different density tissues to create an image. All ultrasound scans were performed by the research fellow using the same ultrasound machine (Siemens Sonoline Antares). This avoided inter-observer variability. I was trained to assess the AFC and ovarian volume by sonographers and clinicians in the Assisted Conception Unit at Guy;s Hospital, where these measurements are performed routinely in women undergoing fertility treatment. Prior to starting the trial, I had 24 training sessions and scanned over 100 patients. I was assessed as competent to locate the ovaries and measure ovarian volume and AFC by a senior sonographer.

In premature ovarian failure it is relatively common to be unable to visualise one or both of the ovaries, due to their small size. In a recent study where experienced fertility specialists assessed the AFC in women with POF, at least one of the ovaries was not visualised in 26% of the women (120). Where it was possible to visualise only one ovary, measurements of AFC and ovarian volume from that ovary were used in the analysis of the mean AFC/ovarian volume.

3.6.1 Antral follicle count

Following identification and magnification of the ovary, a transverse image was obtained and the probe moved very slowly from the superior to inferior aspect. During this movement, all follicles measuring 2-10mm were counted. Antral follicle count (AFC) assessment is widely used in reproductive medicine and the technique has been described previously (121, 123).

Intra-observer variability in AFC assessment has been found to be small (149). The inter-observer variability of AFC measurement is also favourable, and shows a trend towards better reproducibility at lower AFCs (149). It has been suggested that ultrasound AFC measurement is precise enough to be performed as a single measurement by a single observer (149).

The AFCs for four participants were measured first by the research fellow then by a senior sonographer and the values compared. In one of the participants, neither observer visualised the ovaries. In another, neither could view the right ovary. The remaining five ovaries were visualised by both observers and the correlation between their AFC assessments is shown below. When the ovary was not visualised, the AFC was recorded as zero. The correlation coefficient is 0.98 indicating very good agreement between the two observers.

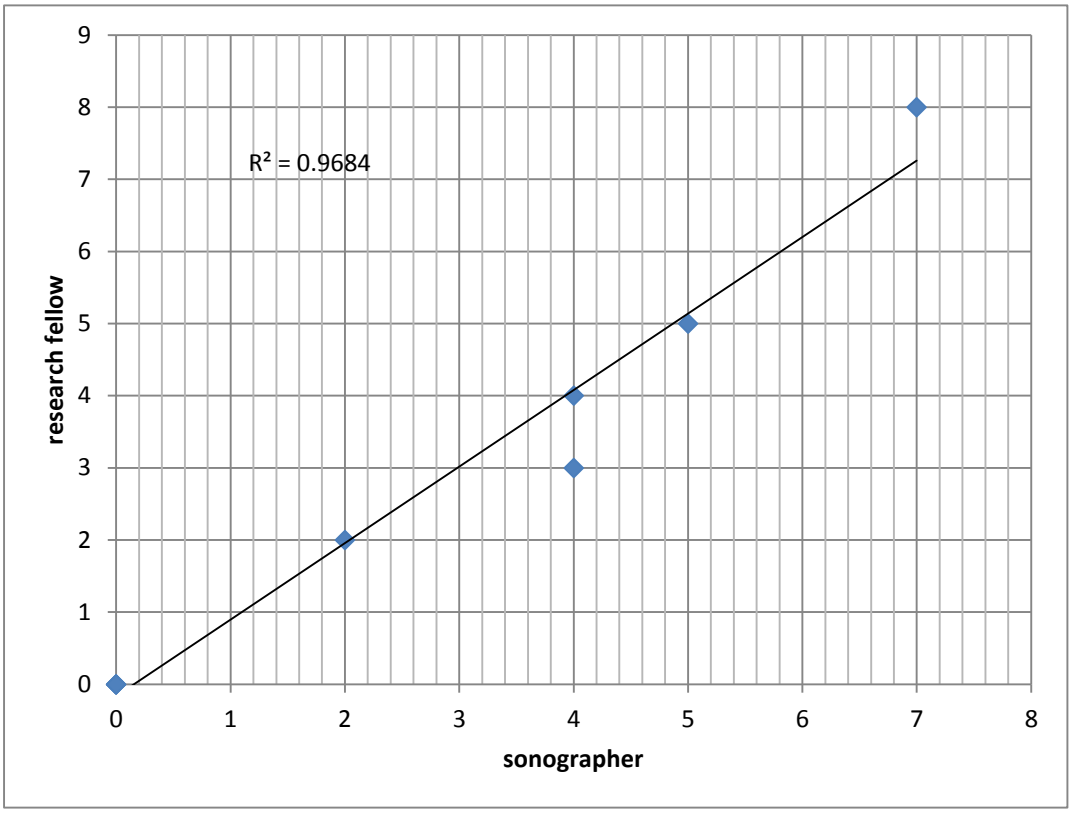


Figure 2 Correlation between antral follicle count measurements

3.6.2 Ovarian volume

The ovarian volume was calculated by obtaining a transverse and longitudinal view of the ovary. Three perpendicular diameters were used to calculate the volume in centimetres cubed.

3.7 Questionnaires

All questionnaires were completed at the start of the visit to avoid bias following interaction with the research fellow. Short, self-administered questionnaires were chosen so that a wide range of relevant areas could be assessed.

3.7.1 Modified Greene Climacteric Scale

The Modified Greene Climacteric Scale (MGCS) (150) gives a brief, overall view of climacteric symptoms. It contains 21 questions encompassing psychological, somatic, vasomotor symptoms, and there is a 'probe question' for sexual dysfunction. The subject is asked to rate the severity of each symptom from none (0) to severe (3). The scores are added to give an overall score, as well as scores for each aspect of the scale. It was developed from factor-analytic studies and was shown to be a reliable test for the assessment of somatic, vasomotor and psychological symptoms (test-retest reliability after 2 weeks 0.83-0.87). Normative data for the general population and menopause clinic are available. It has been recently demonstrated to continue to be a reliable measure of climacteric symptoms (151). The scale has been widely used in clinical trials and Gynaecology clinics to assess the severity of climacteric symptoms and response to treatment (152, 153).

3.7.2 Menopause Symptoms Treatment Satisfaction Questionnaire

The Menopause Symptoms Treatment Satisfaction Questionnaire (MS-TSQ) (154) contains eight questions and aims to assess satisfaction with a study medication over the last four weeks. It is the only validated scale that assesses satisfaction with treatment aimed to alleviate menopausal symptoms. For each question, the participant rates her satisfaction with the study medication from extremely dissatisfied (0) to extremely satisfied (4). A percentage satisfaction score is then calculated. If a single response is missing, the mean of the other responses can be used (154). The threshold of satisfaction of control of hot flushes (calculated using the first two items on the questionnaire – satisfaction with control of hot flushes during day and night) – has been reported as a reduction of 1.6 hot flushes a day (155).

3.7.3 Brief Profile of Female Sexual Function

The Brief Profile of Female Sexual Function (B-PFSF) is a seven item questionnaire with a recall period of 2-3 months. Its aim is to determine the presence of hypo-active sexual desire disorder (HSDD) in post-menopausal women (156). For each item, the subject is asked to rate the frequency of occurrence from never (0) to always (5). The score is added to give a total out of 35. A score of 20 or below indicates HSDD (156).

3.7.4 Patient Health Questionnaire-9

The Patient Health Questionnaire-9 (PHQ-9) has nine questions, which are based on the nine criteria used to make the diagnosis of DSM-IV (Diagnostic and Statistical Manual of Mental Disorders IV) depressive disorder. It aims to detect and assess the severity of depression over the last 2 weeks and can also assess response to treatment (157-159). It is widely used in primary care and reliability has been demonstrated in different ethnic groups (160).

3.7.5 Short Form-36 version 2 (SF-36v2)

The Short Form-36 version 2 (SF-36v2) Health Survey (161) aims to give an objective measure of health-related quality of life. It is the most widely used quality of life tool in clinical trials. Normative data for many populations and age ranges are available. It consists of thirty-six questions with graded responses and provides a 0-100 scale score for each domain. Higher scores indicate better functioning, including in the bodily pain score, where a higher score indicates less limitation due to pain. The SF-36 has a recall period of four weeks and provides scores in the following eight ‘health domain scales’:

- Physical Functioning
- Role-Physical
- Bodily Pain
- General Health
- Vitality
- Social Functioning

- Role-Emotional
- Mental Health

It provides summary scores for physical and mental health (physical component scale and mental component scale). There is also one question on perceived health change over the last year. Scoring is performed by software produced by Quality Metric. The SF-36v2 has been used in one study in iatrogenic POF (56), in which a difference in the Summary of Physical Health score was found between the oestrogen therapy and untreated groups.

3.8 Concordance with medication

Concordance with medication was recorded at each visit as the number of pills left in each empty packet. At the end of the study, the overall percentage of medication taken was calculated. Independent markers of compliance were not performed but unscheduled bleeding was considered to be an adverse event and was enquired about at each visit.

3.9 Calculation of sample size

The initial target sample size was 90 (30 in each group). Assuming that the bone mineral density change within each group varies with a standard deviation of 4% (162, 163), a sample size of 22 in each group would be sufficient to detect a difference in the mean BMD change between groups of 4% assuming a 5% significance level and a power of 90%. The aim was to recruit 30 women to each group in order to make adequate allowance for drop-outs. The following values were used to calculate this:

Estimate of the standard deviation of bone mineral density change in each group: 4%

Significance level: 5%

Difference to be detected: 4%

Power: 90%

Required sample size: 22 in each group (calculated by statistician using Stata)

3.10 Statistical methods

Data were collected on paper Case Record Forms and then entered into a customised Excel spreadsheet. Data entry was checked by Clinical Research Assistants from the King's College London Joint Clinical Trials Office. SPSS (Statistical Package for the Social Sciences) version 19 was used for statistical analysis. We obtained advice from a statistician (Paul Seed) on the best ways to analyse the data. For baseline characteristics, independent sample t-tests were used to make comparisons between the treatment and no treatment groups when the variables were normally distributed, and the Mann-Whitney U test was used in cases of non-parametric distribution (demographics). Baseline characteristics of the HRT and COCP groups were not compared as these groups were randomised.

For the results, all variables were tested for normality using histograms and Q-Q plots. The majority of the variables were found to be normally distributed. For these variables, comparison of changes between the groups were performed using linear regression with adjustment for baseline values. For bone density, comparisons with baseline values were made using paired t-tests, because it is useful clinically to be able to advise a woman how much she can expect her bone density to go up (or down).

The hsCRP and PHQ-9 data were not normally distributed and various transformations including log, inverse log, squaring and square rooting did not succeed in producing a normal distribution. Statistical advice was sought, which was to use the non-parametric Mann-Whitney U test.

Data have been illustrated graphically throughout as both all available data at each time-point and also absolute values for those who completed the trial. This is because due to the high drop-out rate and small numbers, for some variables participants with complete data collection have noticeably different baseline values between groups. It was felt appropriate to illustrate this for clarity.

Chapter 4 Results

4.1 Recruitment

The majority of women were recruited through Guy's and St Thomas' NHS Foundation Trust (GSTFT) – either in the premature ovarian failure clinic or through referrals from colleagues. A significant number were recruited having seen the advertisement on the Daisy Network website. An article on BBC online at the end of November 2010 which included contact details for the Menopause Research Unit led to over 200 enquiries, of which 7 women decided to participate in the study. Table 4 summarises the routes of recruitment.

	Number recruited
GSTFT Premature Ovarian Failure clinic	25
GSTFT Reproductive Medicine clinic	1
Referred from colleague at GSTFT	6
Referred from colleague at another hospital	1
Daisy Network website	11
British Menopause Society website	1
GSTFT intranet and King's College London circular e-mail	5
BBC online article	7
Women's Weekly Advert and Daily Express article	2

Table 4 Numbers recruited through each route. GSTFT – Guy's and St Thomas' NHS Foundation Trust

4.1.1 Recruitment rate

The recruitment rate is shown in figure 3. Recruitment was boosted following a press release in November 2010 which led to BBC online, Daily Express and Women's Weekly articles. The study was also featured on the front page of the hospital intranet.

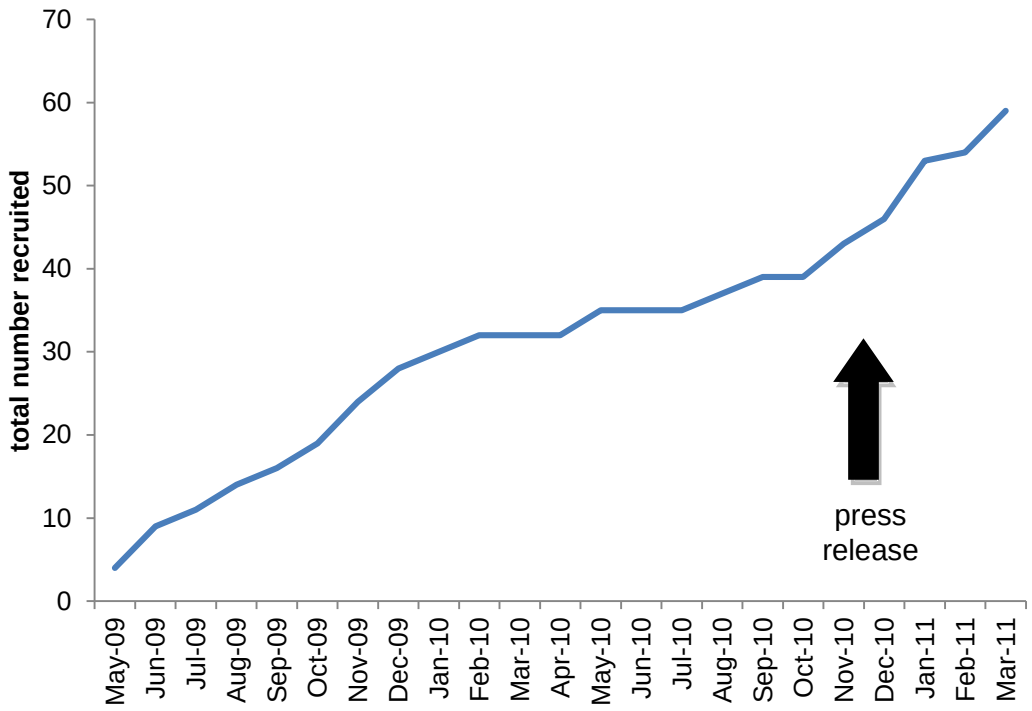


Figure 3 Recruitment rate

4.2 Demographic data

Tables 5 to 7 show the demographic characteristics of study participants in each group.

		HRT	**COCP**	**No treatment**	**All groups**
Age (years)		39.7 (34.2, 43.2)	40.8 (39.5, 42.7)	40.5 (34.5, 43.0)	40.5 (34.8, 42.7)
Ethnicity (%)	White	67	73	59	64
	Mixed	0	0	3	2
	Asian or Asian British	0	0	0	0
	Black or Black British	27	27	38	32
	Chinese or South Asia	7	0	0	2
Occupation (%)	Managerial/professional	33	20	21	24
	Intermediate occupation	47	73	41	51
	Routine/manual	7	7	10	9
	Unemployed	7	0	7	5
	Housewife	7	0	21	12
Current smoking (%)		7	33	10	15
Exercise 3 times a week (%)		53	47	55	53
Alcohol intake (units/week)		8 (1, 10)	1 (0, 3)	2 (0, 10)	2 (0, 10)

Table 5 Sociodemographics. Data shown as mean (SD) or median (25, 75%) as appropriate.

		HRT	COCP	No treatment	All groups
BMI (kg/m^2)		24.7 (3.6)	23.9 (4.6)	24.3 (3.2)	24.4 (3.6)
Systolic blood pressure (mmHg)		122 (11)	117 (8)	113 (14)	121 (12)
Diastolic blood pressure (mmHg)		77 (11)	73 (13)	80 (12)	79 (12)
Family history (%) (where known)	Cardiovascular disease	47	29	32	35
	Hypertension	67	71	61	65
	Breast/ovarian/ endometrial/bowel cancer	47	14	21	26
	Venous thromboembolism	7	14	4	7
Personal history (%)	Cardiovascular disease	0	0	0	0
	Hypercholesterolaemia	0	0	0	0
	Hypertension	7	0	3	3
	Breast cancer	0	0	0	0
	Migraine	13	13	35	24
	Venous	0	0	0	0
	thromboembolism	7	7	10	9
	Thyroid disease	0	0	0	0
	Diabetes	0	7	0	2
	Vitiligo				

Table 6 General risk factors. Data shown as mean (SD) or median (25, 75%) as appropriate.

		HRT	COCP	No treatment	All groups
Cause of POF (%)	No cause identified Autoimmune FMR1 premutation	93 7 0	87 13 0	86 10 3	88 10 2
Family history POF (%)		27	28	25	26
Mother's age at menopause (%)	Under 45 45 or older Not known	20 60 20	20 53 27	10 35 55	15 46 39
Time since diagnosis POF (months)		5 (2, 23)	21 (9, 32)	11 (5, 28)	9 (5, 28)
Time since LMP (months)		11 (3, 25)	13 (5, 29)	3 (1, 22)	6 (2, 25)
Previous use of HRT/COCP for POF (%)		33	60	14	31
Previous use of COCP for other reasons (%)		67	60	72	68
Any previous pregnancy (%)		73	53	62	63
Number of children	0 1 2 3 4	53 33 7 7 0	60 13 20 7 0	59 10 10 14 7	58 17 12 10 3
Current desire for pregnancy (including those with no partner) (%)		7	0	41	22

Table 7 Factors relevant to POF. Data shown as mean (SD) or median (25, 75%) as appropriate

4.3 Baseline symptoms

Table 8 shows the baseline symptoms in each group.

	HRT	COCP	No treatment	All groups
Hot flushes	60	73	38	52
Night sweats	33	53	45	44
Sleep disturbance	67	53	45	53
Change in mood	67	67	66	66
Lack of energy	53	67	55	58
Vaginal dryness	47	53	35	42
Decreased libido	60	67	48	60
Dyspareunia	27	33	25	23
Palpitations	47	47	17	32
Lack of concentration	53	60	45	51
Urinary problems	27	27	24	25
Headache	47	33	35	37
Skin itching	33	40	14	25
Joint pain	47	53	38	44
Depression	20	53	35	36
Anxiety	47	40	41	42
Irritability	60	60	55	58
Low self esteem	33	47	31	36
Other	33	20	17	22

Table 8 Symptoms at baseline (%)

Other symptoms included dry skin 3%, hair thinning 3%, fainting/dizziness 3%, leg cramps 2%, tinnitus 2%, increased facial hair 3%, weight gain 2%, stiff joints 2% and red patches on skin 2%.

4.4 Screen failures

Five women were not eligible for the trial following screening. One had an initial FSH of below 30IU/l and the other four had a second FSH value of below 30IU/l. There were no other reasons for screen failure.

4.5 Reasons behind choosing the treatment or no treatment group

59 women were recruited to the study – 29 to the no treatment group and 30 to the treatment group. The most common reason cited for declining treatment was a dislike of taking medication. The main reason for choosing the treatment group was menopausal symptoms, followed by concerns about osteoporosis and cardiovascular disease. Figures 4 and 5 show the main reasons women cited for their choices.

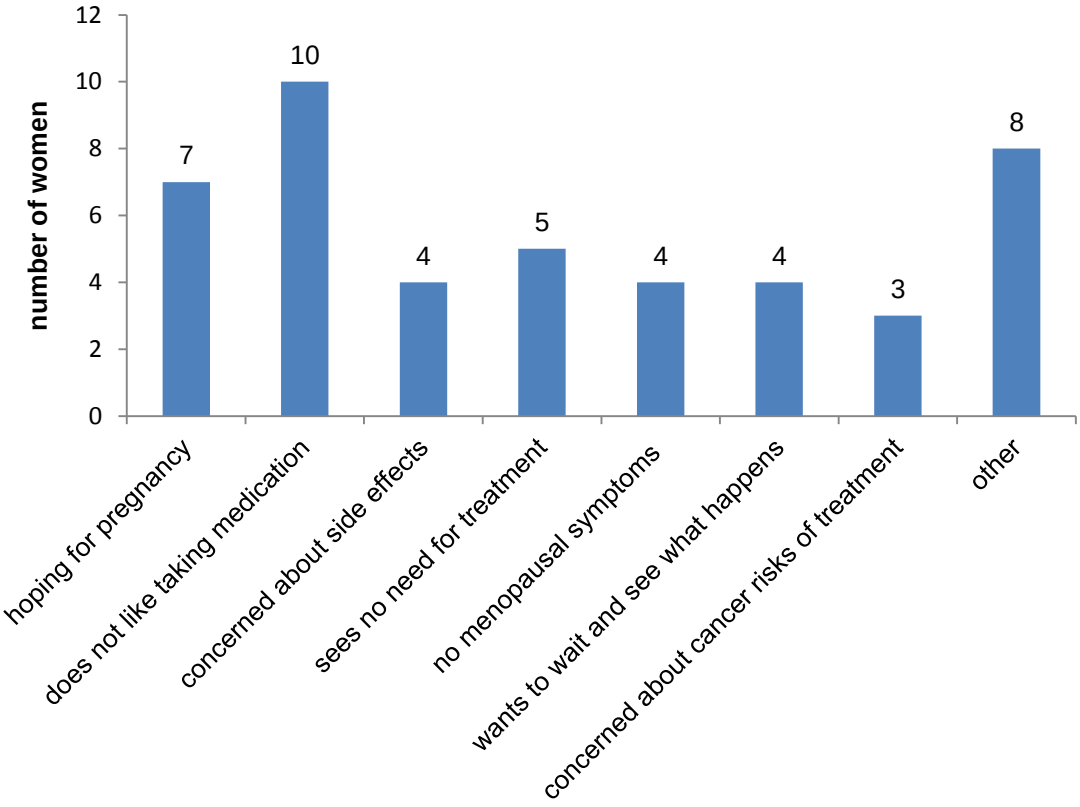


Figure 4 Reasons for choosing no treatment. 'Other' reasons included a mistrust of medication, themes of wanting to 'stay natural' or not put chemicals into the body and concerns over controversial press regarding HRT. This was following extensive counselling on the benefits of oestrogen replacement.

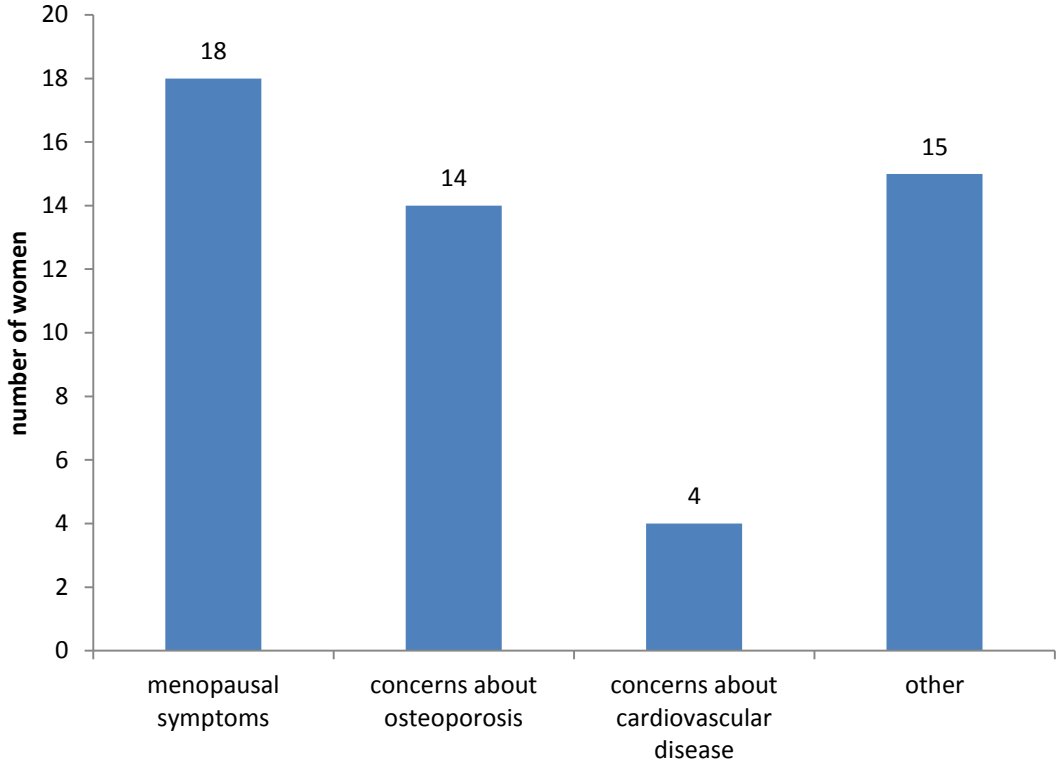


Figure 5 Reasons for choosing treatment. 'Other' reasons included wanting to get hormones the body needs; feeling that body is lacking something; to keep tissues healthy; to treat symptoms of dry skin/fatigue; wanting to see if it makes any difference

4.6 Baseline comparisons between the treatment and no treatment groups

Independent sample t-tests were used to compare baseline characteristics between the treatment and no treatment groups when the variables were normally distributed, and the Mann-Whitney U test was used in cases of non-parametric distribution (demographics). p-values of <0.05 were seen between treatment and no treatment groups for Modified Greene Climacteric Scale (MGCS), Short Form-36 (SF-36) mental component score and Patient Health Questionnaire-9 (PHQ-9) score. The treatment groups had a higher MGCS, indicating more menopausal symptoms, and had a higher depressions score (PHQ-9) and lower mental health summary score (SF-36), both indicating poorer mental health. Baseline characteristics of the HRT and COCP groups were not compared as these groups were randomised.

	HRT	COCP	All treatment (HRT and COCP)	No treatment	p-value (treatment vs no treatment)
Demographics					
Age (years)	39.7 (34.2, 43.2)	40.8 (39.5, 42.7)	38.8 (37.5, 42.8)	40.5 (34.5, 43.0)	0.688
Months since diagnosis POF	5 (2, 23)	21 (9, 32)	9 (4, 28)	11 (5, 28)	0.611
Months since LMP	11 (3, 25)	13 (5, 29)	12 (5, 25)	3 (1, 22)	0.063
Blood pressure and BMI					
Systolic bp (mmHg)	122 (11)	117 (8)	119 (10)	113 (14)	0.256
Diastolic bp (mmHg)	77 (11)	73 (13)	77 (12)	80 (12)	0.271
BMI (kg/m^2)	24.7 (3.6)	23.9 (4.6)	25 (4)	24.3 (3.2)	0.771
Questionnaires					
MGCS score (total)	22 (17)	24 (14)	23 (15)	16 (12)	**0.037**
BPFSF score	18 (7)	18 (9)	18 (8)	20 (8)	0.305
SF-36 physical component score	52 (9)	50 (8)	51 (8)	52 (8)	0.513
SF-36 mental component score	43 (14)	37 (16)	40 (15)	47 (10)	**0.040**
PHQ-9 score	9 (7)	10 (7)	9 (7)	6 (5)	**0.028**
Bone mineral density and Z scores					
Spine Z-score	-0.6 (1.3)	-0.5 (1.3)	-0.52 (1.26)	-0.62 (1.23)	0.750
Spine BMD (g/cm^2)	0.985 (0.133)	0.996 (0.153)	0.991 (0.141)	0.988 (0.119)	0.928
Total hip Z-score	-0.1 (1.0)	-0.5 (0.9)	-0.28 (0.95)	-0.155 (0.84)	0.596
Total hip BMD (g/cm^2)	0.929 (0.121)	0.894 (0.153)	0.912 (0.136)	0.929 (0.098)	0.584

Table 9 Baseline comparisons between treatment and no treatment groups. BMI body mass index, MGCS Modified Greene Climacteric Scale, SF-36 Short Form 36, BMD bone mineral density, bp blood pressure. Data shown as mean (SD) or median (25, 75%) as appropriate.

4.7 Withdrawals from the study

36 out of 59 women completed the study (61%). The percentage completing follow up in each group was 52% in the no treatment group, 80% in the HRT group and 60% in the COCP group. The timing of drop-outs and reasons is shown in the flowchart.

The most frequent reason for withdrawal was loss to follow up. There was a slightly higher drop-out rate in the COCP group compared with the HRT group, but this was due to loss of follow up. One participant in each of the treatment groups withdrew because she felt that the medication was contributing to symptoms of depression and she wanted to change medication. One participant in each treatment group withdrew due to side effects – in the COCP group this was breast tenderness and in the HRT group the participant reported that the medication was not helping, was making her symptoms worse, causing a headache and making her feel 'not myself'. One participant in the no treatment group withdrew for two reasons – a drop in bone density and also menopausal symptoms.

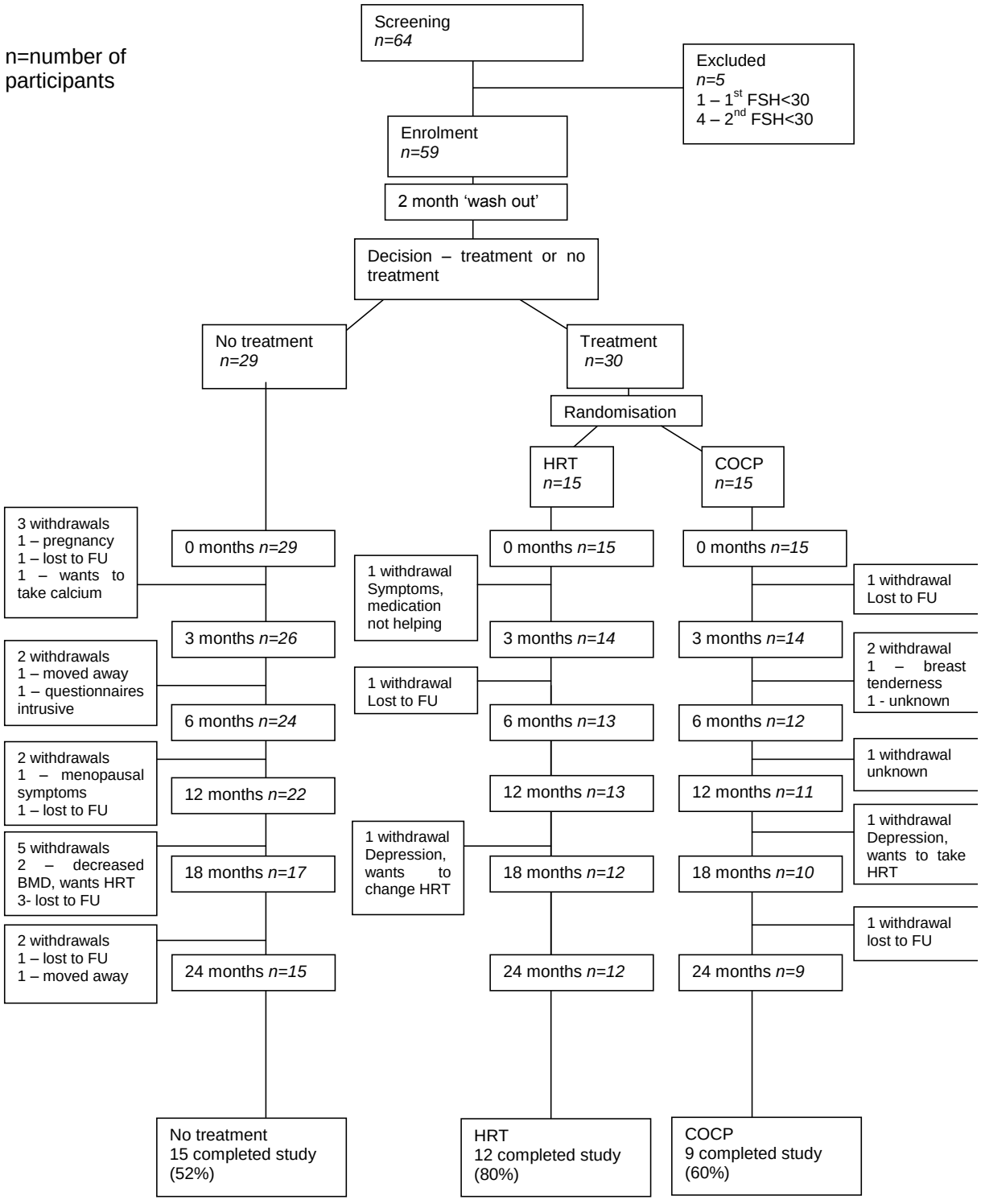


Figure 6 Trial flowchart

4.8 Compliance with medication

Compliance with medication in the two treatment groups was calculated from the number of pills left in empty packets at the 6, 12, 18 and 24 month study visits. On the few occasions when this was not recorded, the percentages from the visits where it was recorded are used. For those who completed the trial, compliance was 95% in the COCP and 98% in the HRT group. Overall, compliance was 95% in the COCP and 98% in the HRT group. We did not perform independent markers of compliance.

4.9 Other medications taken during the study

29 women took either prescribed or over the counter medication during the course of the study. Three women took antidepressants; one woman in the HRT group was already taking this at the start of the study and one woman in each of the other groups started an antidepressant during the study. Three women took antihypertensives; one woman in the HRT group was already taking an antihypertensive at the start of the trial and one woman in each of the other groups started during the study. Two women were already taking thyroxine at the start of the trial (one in the HRT and one in the COCP group) and continued on it throughout. Eleven women took analgesia during the study. Four women took medication for asthma – three in the HRT group were taking this at the start of the study and one woman in the no treatment group started during the study.

Other medications taken were ferrous sulphate (two women), enoxaparin, a short course of prednisolone, loperamide, omeprazole and general multivitamins.

4.10 Adverse events

39 women in the study experienced at least one adverse event (17 in the no treatment group, 11 in the HRT group and 11 in the COCP group). There was one serious adverse event, which was a pregnancy in the no treatment group. This resulted in the birth of a healthy baby.

The most common adverse events are detailed in table 10 below.

	HRT	COCP	No treatment
Breast tenderness	2	3	0
Breast lump	1	0	1
Pelvic pain/dysmenorrhoea	3	2	1
Hot flushes /night sweats	4	2	6
Headache/migraine	1	2	1
Depression/feeling low	1	1	2
Musculo-skeletal pain	0	2	3
Pregnancy	0	0	1
Nausea	0	1	0
Infection (chest/upper respiratory tract/ear)	1	1	3
Urinary tract infection	1	1	0
Chest pains	0	1	1

Table 10 Numbers of most common adverse events by group

In addition, in the no treatment group one woman had carpal tunnel syndrome which required decompression, one woman had anaemia which required investigation by endoscopy and there was one each of the following adverse events: dysuria, umbilical hernia, gastric reflux and fall.

In the HRT group, one woman experienced patches of brown skin on her face, one had a simple ovarian cyst which required monitoring and one had an allergic skin reaction (which had happened to her previously when not on HRT). Other adverse events in the HRT group were facial acne, worsening of menopausal symptoms, feeling edgy and nervous, longer periods and a fractured ankle due to trauma.

In the COCP group, one woman experienced deterioration in eyesight and was under investigation for this by an ophthalmologist. One woman had irregular vaginal spotting, one suffered from insomnia and one complained of a dry mouth. Other adverse events reported in

the COCP group were increased appetite and impaired glucose tolerance, each in a single participant.

The breast lump and chest pains were investigated and found to be benign.

4.11 Bone Mineral Density

Comparisons with baseline values are shown in these results because it is useful clinically to be able to advise a patient how much bone loss can be anticipated in a given situation and how this may be affected by treatment. Comparisons with baseline values are not used in the rest of the results because the main aim of the trial is to compare treatments. Paired t-tests were used to compare results at 6, 12 and 24 months with baseline bone mineral density (BMD).

At the lumbar spine, there was a significant gain in BMD with HRT at all time points, a drop in BMD in the no treatment group at 12 and 24 months and no change in BMD in the COCP group. At the total hip, bone density was maintained in the HRT and COCP groups but did not increase over 24 months, whereas in the no treatment group there was a significant drop at all time points (figs 7 to 12). At the femoral neck, there was a reduction in bone density in the no treatment group over the course of the trial, although this only became significant at 24 months. There were no significant changes compared with baseline values in the HRT or COCP groups.

Comparison between the groups was performed using linear regression with adjustment for baseline bone mineral density. This revealed a significant difference between the COCP and HRT groups in lumbar spine BMD at 12 and 24 months, in favour of HRT. Fig 8 shows that women with complete data collection had comparable baseline values. (This was not the case in some other results, particularly in the questionnaires.) There were no differences between the COCP and HRT groups in change in total hip BMD. At the femoral neck, there was a significant difference at 12 months in favour of HRT.

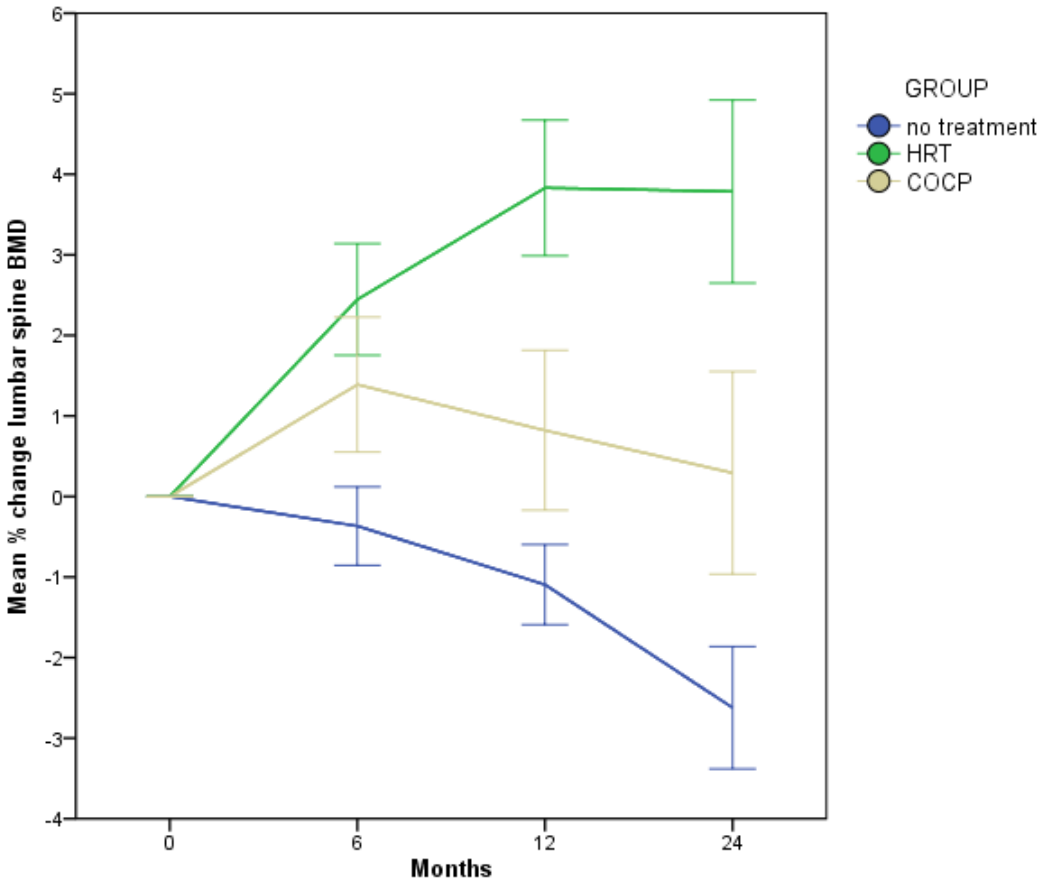
Comparison between the HRT and no treatment groups of changes at both the lumbar spine and the hip were highly significant at all time-points in favour of HRT. At the femoral neck the differences were significant at 12 and 24 months.

Comparison of the COCP and no treatment groups at the lumbar spine revealed a trend in favour of COCP, but no statistically significant differences were found and a larger sample size would be needed to confirm any real difference. However, the total hip results were similar to the HRT/no treatment comparison, with highly significant differences between the COCP and no treatment groups at all time-points in favour of COCP. No significant differences were found between the COCP and no treatment groups at the femoral neck.

4.11.1 Lumbar spine bone mineral density

	HRT		**COCP**		**No treatment**	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline lumbar spine BMD (g/m^2)	15	0.985 (0.133)	15	0.996 (0.153)	29	0.988 (0.119)
Change from baseline to 6 months (g/m^2)	10	+0.024 (0.021)	11	+0.011 (0.024)	21	-0.003 (0.022)
% change from baseline to 6 months		+2.45		+1.39		-0.37
p value 6 months vs baseline BMD		0.005		0.145		0.476
Change from baseline to 12 months (g/m^2)	13	+0.038 (0.028)	11	+0.006 (0.032)	20	-0.011 (0.022)
% change from baseline to 12 months		+3.83		+0.82		-1.09
p value 12 months vs baseline BMD		<0.001		0.528		0.042
Change from baseline to 24 months (g/m^2)	12	+0.039 (0.037)	9	0.000 (0.039)	15	-0.027 (0.031)
% change from baseline to 24 months		+3.79		+0.29		-2.62
p value 24 months vs baseline BMD		0.004		0.987		0.005

Table 11 Bone mineral density (BMD) results at the lumbar spine (g/m^2) and comparisons with baseline



COCP vs HRT at 24 months p=0.040

Figure 7 Percent changes from baseline in mean lumbar spine bone mineral density (BMD) over 24 months showing all available data at each time-point. Data shown as mean +/- 1 standard error.

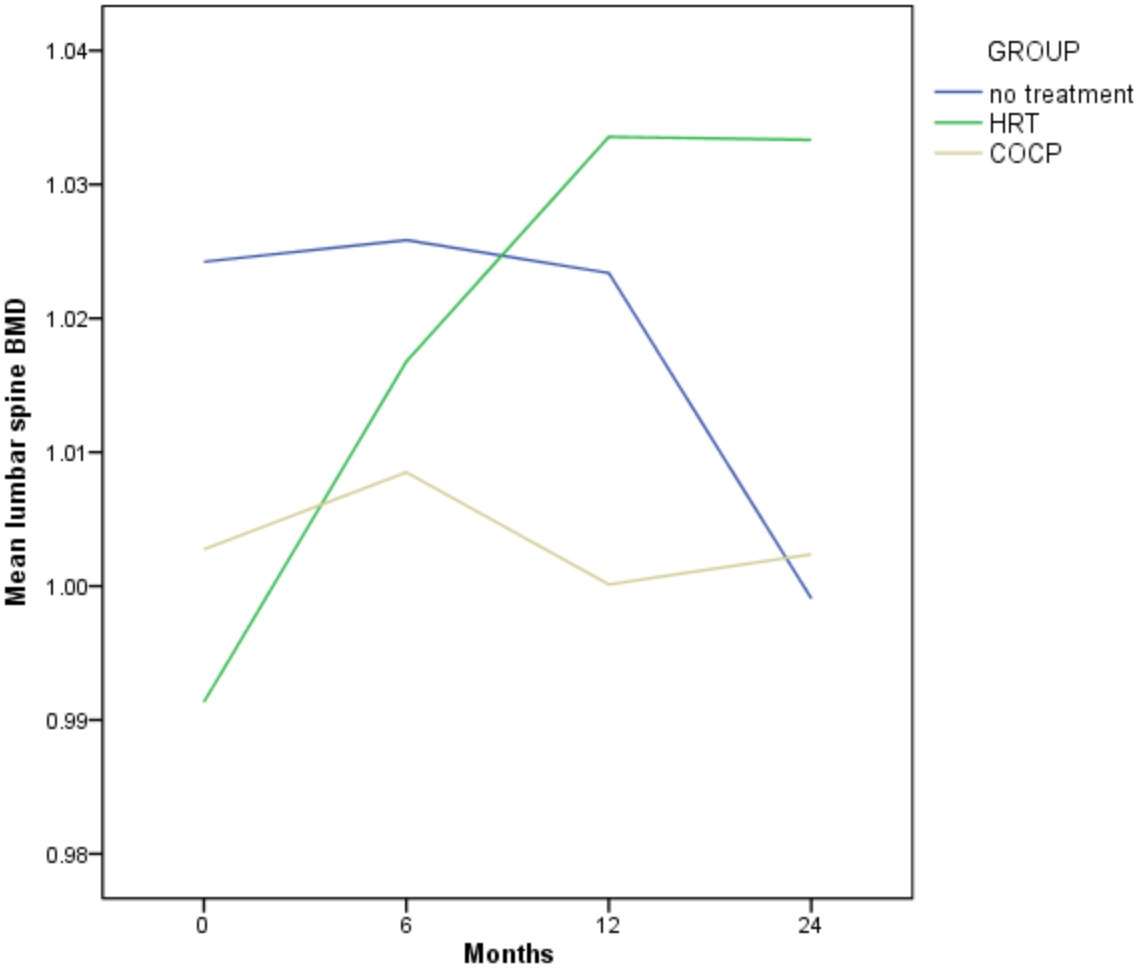


Figure 8 Changes in mean lumbar spine bone mineral density (BMD) (g/m^2) over 24 months in participants with complete data collection

4.11.1.1 Comparison between COCP and HRT

Months	COCP minus HRT mean lumbar spine BMD	95% confidence interval of the difference	p value
6	-0.013	-0.034 to 0.008	0.216
12	-0.032	-0.058 to -0.005	0.021
24	-0.038	-0.073 to -0.002	0.040

Table 12 Comparison between HRT and COCP lumbar spine bone mineral density (BMD) (g/m^2) results. Linear regression analysis was used to adjust for baseline score

4.11.1.2 Comparison between HRT and no treatment

Months	HRT minus no treatment mean lumbar spine BMD	95% confidence interval of the difference	p value
6	0.028	0.011 to 0.045	0.002
12	0.049	0.031 to 0.068	<0.001
24	0.065	0.038 to 0.093	<0.001

Table 13 Comparison between HRT and no treatment lumbar spine bone mineral density (BMD) (g/m^2) results. Linear regression analysis was used to adjust for baseline score

4.11.1.3 Comparison between COCP and no treatment

Months	COCP minus no treatment mean lumbar spine BMD	95% confidence interval of the difference	p value
6	0.007	-0.001 to 0.016	0.099
12	0.009	-0.001 to 0.019	0.091
24	0.014	-0.001 to 0.028	0.066

Table 14 Comparison between COCP and no treatment lumbar spine bone mineral density (BMD) (g/m^2) results. Linear regression analysis was used to adjust for baseline score

4.11.2 Total hip bone mineral density

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline total hip BMD (g/m^2)	15	0.929 (0.120)	15	0.894 (0.152)	29	0.929 (0.098)
Change from baseline to 6 months (g/m^2)	10	+0.003 (0.015)	11	+0.012 (0.016)	21	-0.011 (0.017)
% change from baseline to 6 months		+0.32		+1.41		-1.17
p value 6 months vs baseline BMD		0.540		0.033		0.007
Change from baseline to 12 months (g/m^2)	13	0.007 (0.016)	11	0.007 (0.018)	20	-0.014 (0.019)
% change from baseline to 12 months		+0.75		+0.71		-1.52
p value 12 months vs baseline BMD		0.154		0.232		0.005
Change from baseline to 24 months (g/m^2)	12	0.007 (0.022)	9	0.002 (0.022)	15	-0.023 (0.015)
% change from baseline to 24 months		+0.83		+0.33		-2.48
p value 24 months vs baseline BMD		0.295		0.803		<0.001

Table 15 Bone mineral density (BMD) results at the hip (g/m^2) and comparisons with baseline

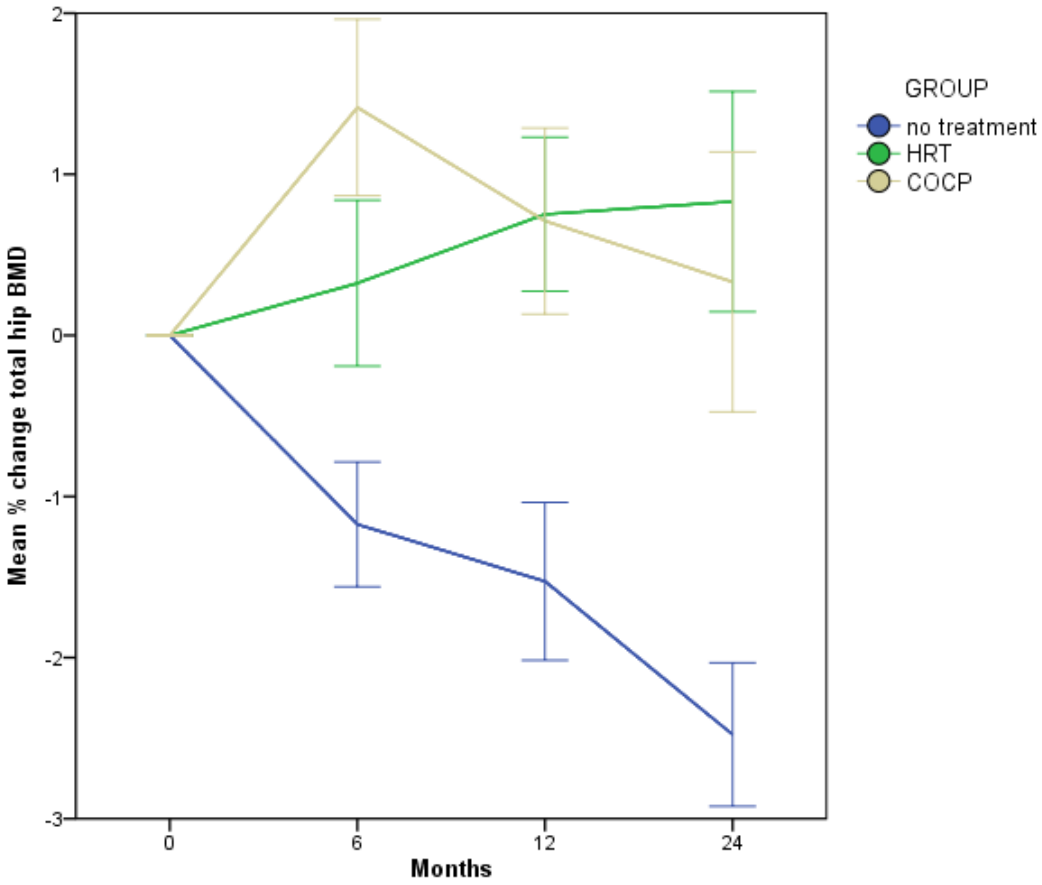


Figure 9 Percent changes from baseline in mean total hip bone mineral density (BMD) over 24 months showing all available data at each time-point. Data shown as mean +/- 1 standard error.

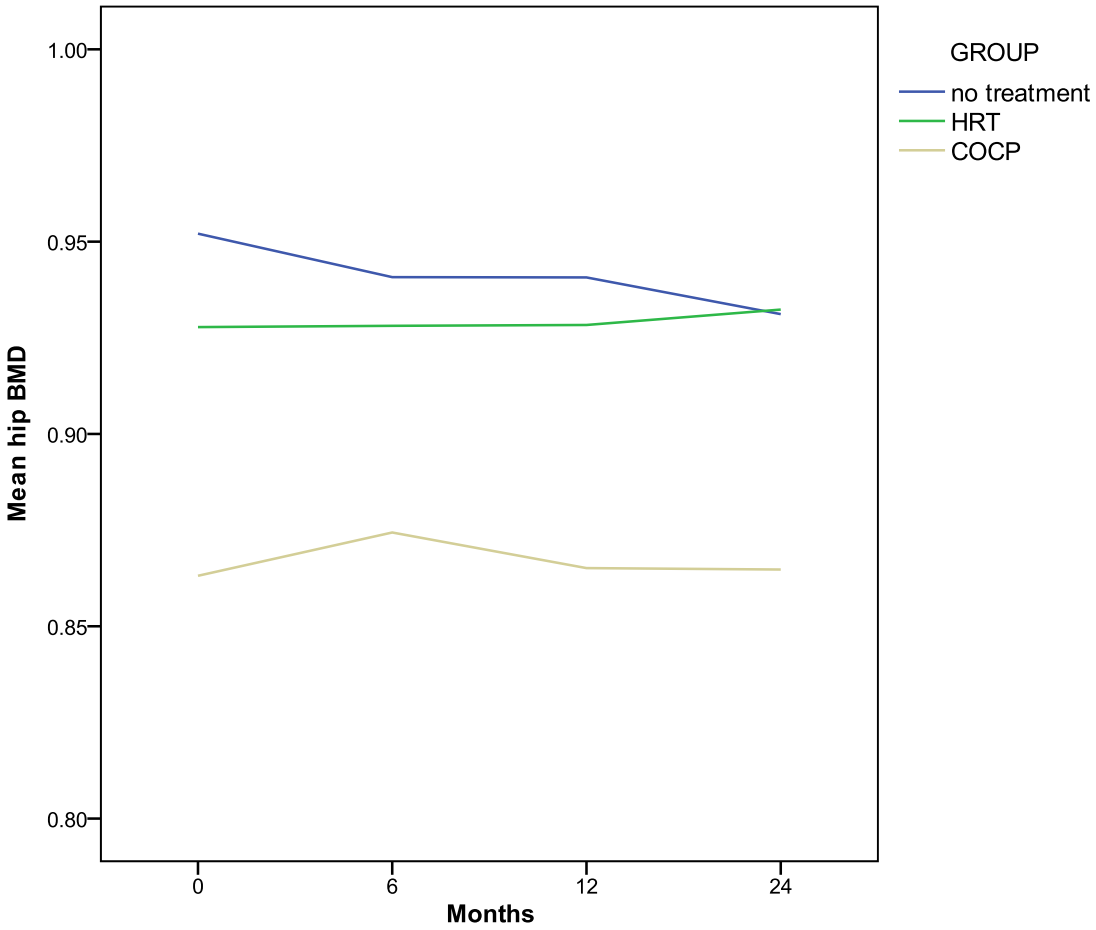


Figure 10 Changes in mean total hip bone mineral density (BMD) (g/m^2) over 24 months in participants with complete data collection

4.11.2.1 Comparison between HRT and COCP

Months	COCP minus HRT mean total hip BMD	95% confidence interval of the difference	p value
6	0.008	-0.007 to 0.023	0.294
12	0.000	-0.015 to 0.015	0.976
24	-0.008	-0.027 to 0.012	0.431

Table 16 Comparison between HRT and COCP total hip bone mineral density (BMD) (g/m^2) results. Linear regression analysis was used to adjust for baseline score

4.11.2.2 Comparison between HRT and no treatment

Months	HRT minus no treatment mean total hip BMD	95% confidence interval of the difference	p value
6	0.014	0.001 to 0.027	0.036
12	0.020	0.007 to 0.034	0.004
24	0.030	0.015 to 0.044	<0.001

Table 17 Comparison between HRT and no treatment total hip bone mineral density (BMD) (g/m^2) results. Linear regression analysis was used to adjust for baseline score

4.11.2.3 Comparison between COCP and no treatment

Months	COCP minus no treatment mean total hip BMD	95% confidence interval of the difference	p value
6	0.011	0.004 to 0.018	0.002
12	0.011	0.004 to 0.018	0.005
24	0.012	0.004 to 0.020	0.005

Table 18 Comparison between COCP and no treatment total hip bone mineral density (BMD) (g/m^2) results. Linear regression analysis was used to adjust for baseline score

4.11.3 Femoral neck bone mineral density

	HRT		**COCP**		**No treatment**	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline femoral neck BMD (g/m^2)	15	0.813 (0.113)	15	0.761 (0.118)	29	0.816 (0.106)
Change from baseline to 6 months (g/m^2)	10	+0.013 (0.019)	11	+0.015 (0.039)	21	-0.004 (0.034)
% change from baseline to 6 months		+1.59		+1.59		-0.24
p value 6 months vs baseline BMD		0.069		0.243		0.644
Change from baseline to 12 months (g/m^2)	13	+0.013 (0.031)	11	-0.013 (0.019)	20	-0.007 (0.022)
% change from baseline to 12 months		+1.66		-1.61		-0.91
p value 12 months vs baseline BMD		0.161		0.071		0.175
Change from baseline to 24 months (g/m^2)	12	+0.009 (0.020)	9	+0.002 (0.027)	15	-0.014 (0.025)
% change from baseline to 24 months		+1.16		+0.42		-1.78
p value 24 months vs baseline BMD		0.146		0.858		0.048

Table 19 Bone mineral density (BMD) results at the femoral neck (g/m^2) and comparisons with baseline

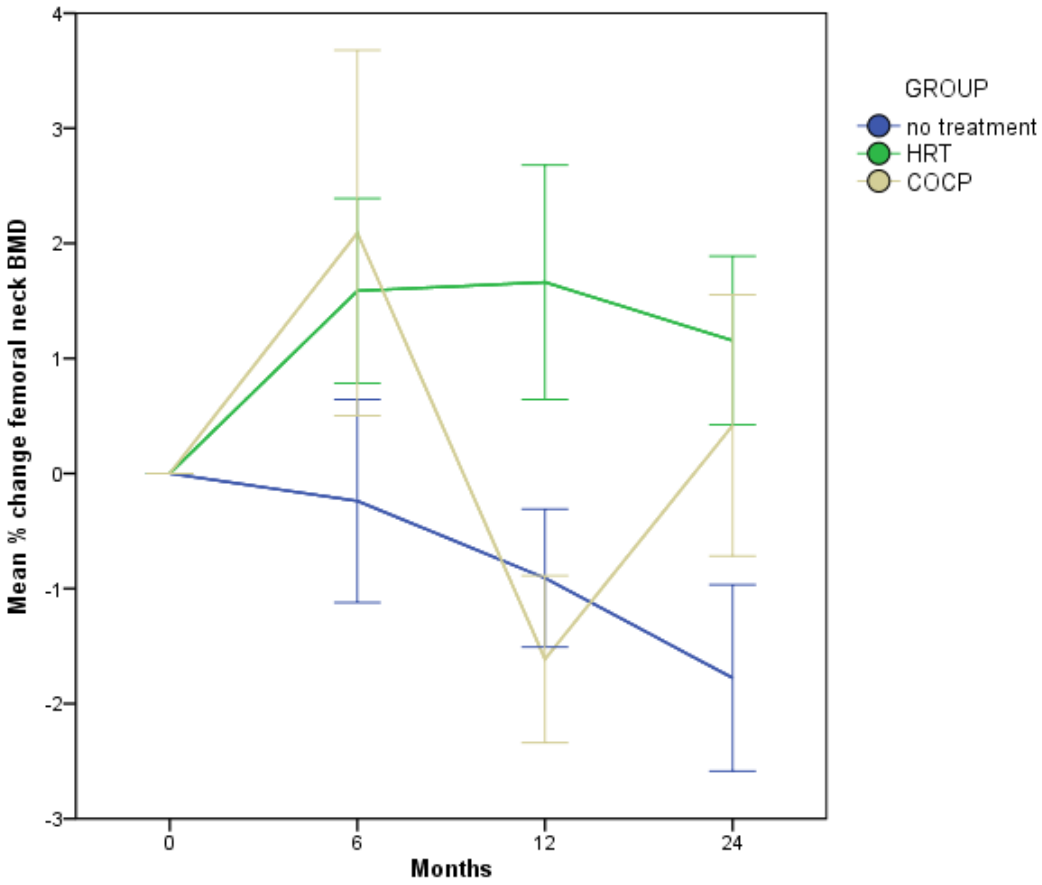


Figure 11 Percent changes from baseline in mean femoral neck bone mineral density (BMD) over 24 months showing all available data at each time-point. Data shown as mean +/- 1 standard error.

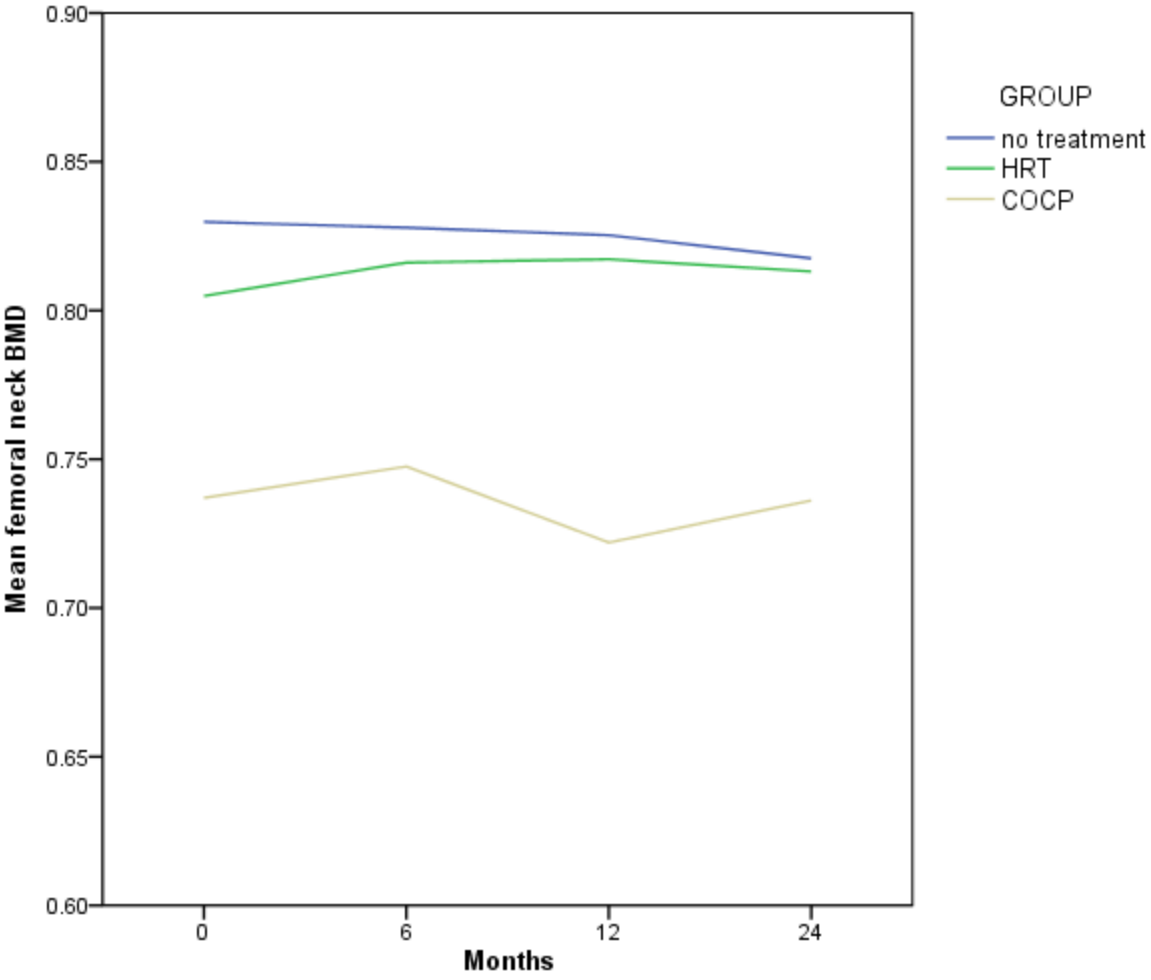


Figure 12 Changes in mean femoral neck bone mineral density (BMD) (g/m^2) over 24 months in participants with complete data collection

4.11.3.1 Comparison between COCP and HRT

Months	COCP minus HRT mean femoral neck BMD (g/m^2)	95% confidence interval of the difference (g/m^2)	p value
6	0.000	-0.030 to 0.030	0.980
12	-0.027	-0.052 to -0.002	0.033
24	-0.009	-0.031 to 0.012	0.378

Table 20 Comparison between HRT and COCP femoral neck bone mineral density (BMD) (g/m^2) results. Linear regression analysis was used to adjust for baseline score

4.11.3.2 Comparison between HRT and no treatment

Months	HRT minus no treatment mean femoral neck BMD (g/m^2)	95% confidence interval of the difference (g/m^2)	p value
6	0.015	-0.009 to 0.038	0.212
12	0.020	0.001 to 0.040	0.043
24	0.023	0.005 to 0.042	0.017

Table 21 Comparison between HRT and no treatment femoral neck bone mineral density (BMD) (g/m^2) results. Linear regression analysis was used to adjust for baseline score

4.11.3.3 Comparison between COCP and no treatment

Months	COCP minus no treatment mean femoral neck BMD (g/m^2)	95% confidence interval of the difference (g/m^2)	p value
6	0.005	-0.009 to 0.019	0.474
12	-0.003	-0.012 to 0.006	0.548
24	0.008	-0.004 to 0.020	0.185

Table 22 Comparison between COCP and no treatment femoral neck bone mineral density (BMD) (g/m^2) results. Linear regression analysis was used to adjust for baseline score

4.12 Bone turnover markers

Both procollagen type I N-terminal propeptide (P1NP) and C-terminal cross-linked telopeptide (CTX) were reduced from baseline at all time points in the HRT and COCP groups. The graphs show a trend towards a greater reduction in P1NP in the HRT group, but on statistical comparison there were no significant differences between the HRT and COCP groups on the extent of reduction of either CTX or P1NP. In the no treatment group there was a slight increase in P1NP and CTX over the course of the trial. When each treatment group was compared with the no treatment group, the differences were significant at every time-point for both CTX and P1NP. The graphs showing women with complete data collection indicate that baseline values were very similar between the groups for both P1NP and CTX.

4.12.1 Procollagen type I N-terminal propeptide

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline P1NP	15	67.7 (24.7)	15	56.1 (31.9)	29	59.6 (21.7)
Change from baseline to 6 months	11	-27.5 (26.3)	12	-20.6 (29.8)	23	+3.0 (17.3)
Change from baseline to 12 months	13	-26.3 (14.5)	10	-17.1 (35.2)	21	+5.8 (29.7)
Change from baseline to 24 months	12	-23.4 (12.8)	9	-15.4 (38.4)	15	+6.8 (24.8)

Table 23 Procollagen type I N-terminal propeptide (P1NP) results (mcg/l)

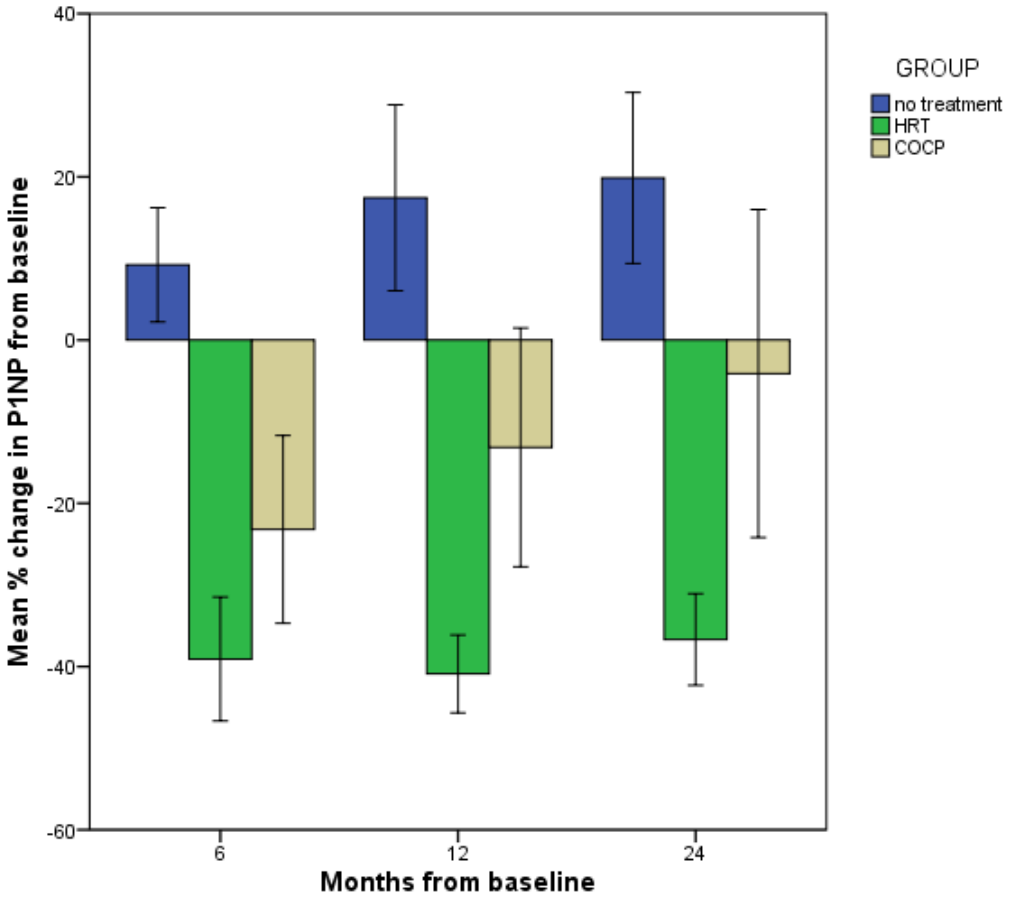


Figure 13 Percent changes from baseline in procollagen type I N-terminal propeptide (P1NP) results over 24 months showing all available data at each time-point. Data shown as mean +/- one standard error.

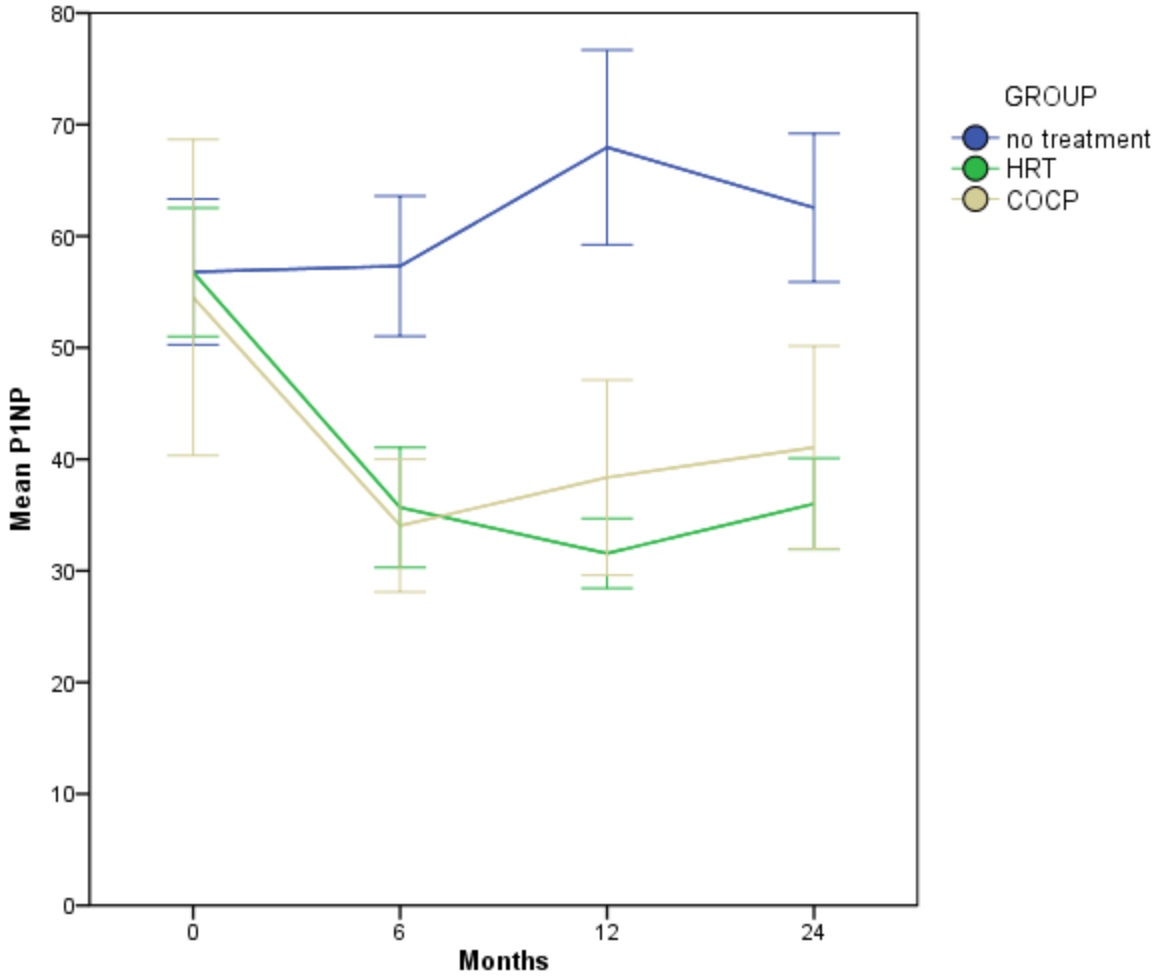


Figure 14 Changes in mean procollagen type I N-terminal propeptide (P1NP) results (mcg/l) over 24 months in participants with complete data collection. Data shown as mean +/- 1 standard error.

4.12.1.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean P1NP	95% confidence interval of the difference	p value
6	3.5	-11.8 to 18.7	0.638
12	7.2	-6.5 to 20.9	0.288
24	5.8	-10.1 to 21.6	0.455

Table 24 Comparison between HRT and COCP procollagen type I N-terminal propeptide (P1NP) results (mcg/l). Linear regression analysis was used to adjust for baseline score

4.12.1.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean P1NP	95% confidence interval of the difference	p value
6	-29.4	-42.6 to -16.1	<0.001
12	-30.4	-46.9 to -13.9	0.001
24	-28.4	-42.3 to -14.5	<0.001

Table 25 Comparison between HRT and no treatment procollagen type I N-terminal propeptide (P1NP) results (mcg/l). Linear regression analysis was used to adjust for baseline score

4.12.1.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean P1NP	95% confidence interval of the difference	p value
6	-12.3	-18.8 to -5.9	<0.001
12	-11.4	-21.7 to -1.0	0.032
24	-11.0	-21.0 to -0.1	0.033

Table 26 Comparison between COCP and no treatment procollagen type I N-terminal propeptide (P1NP) results (mcg/l). Linear regression analysis was used to adjust for baseline score

4.12.2 C-terminal cross-linked telopeptide

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline CTX	15	0.41 (0.17)	15	0.39 (0.18)	29	0.38 (0.18)
Change from baseline to 6 months	11	-0.20 (0.20)	12	-0.17 (0.14)	23	-0.01 (0.20)
Change from baseline to 12 months	13	-0.18 (0.12)	10	-0.20 (0.16)	21	0.03 (0.15)
Change from baseline to 24 months	12	-0.11 (0.13)	9	-0.08 (0.20)	15	0.06 (0.11)

Table 27 C-terminal cross-linked telopeptide (CTX) results (mcg/l)

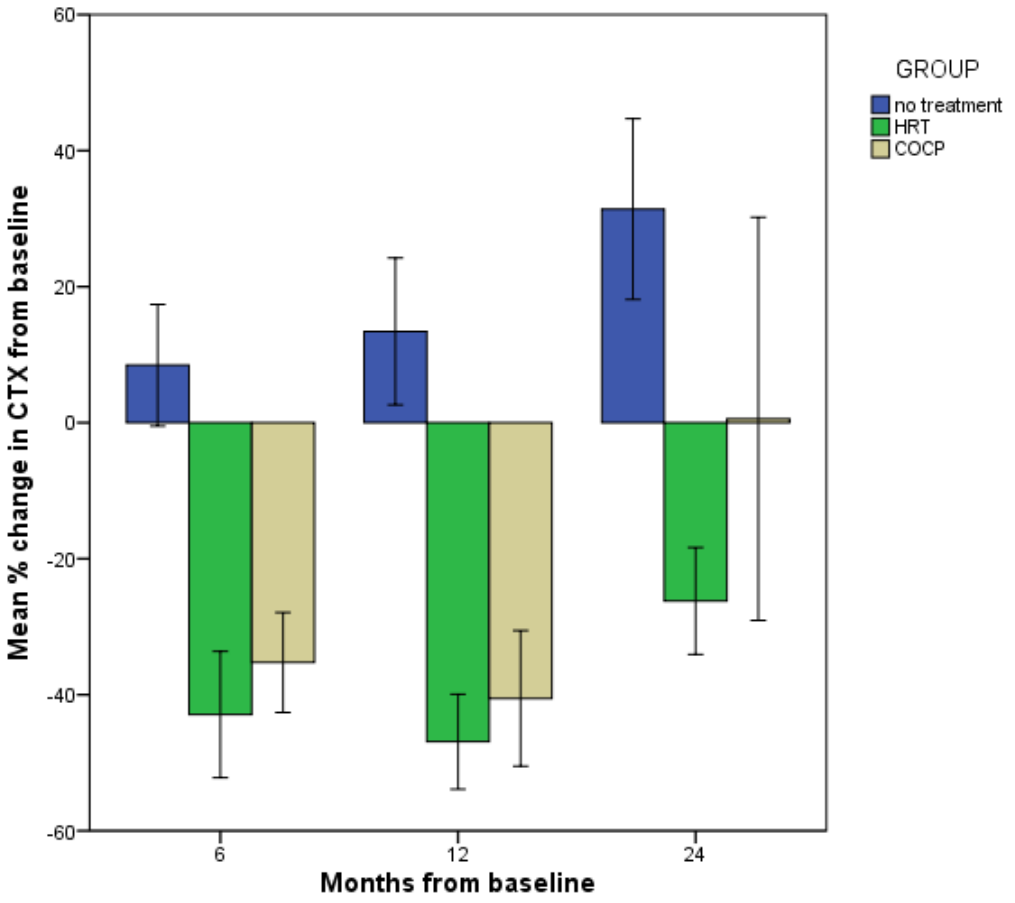


Figure 15 Percent changes from baseline in C-terminal cross-linked telopeptide (CTX) results over 24 months showing all available data at each time-point. Data shown as mean +/- one standard error.

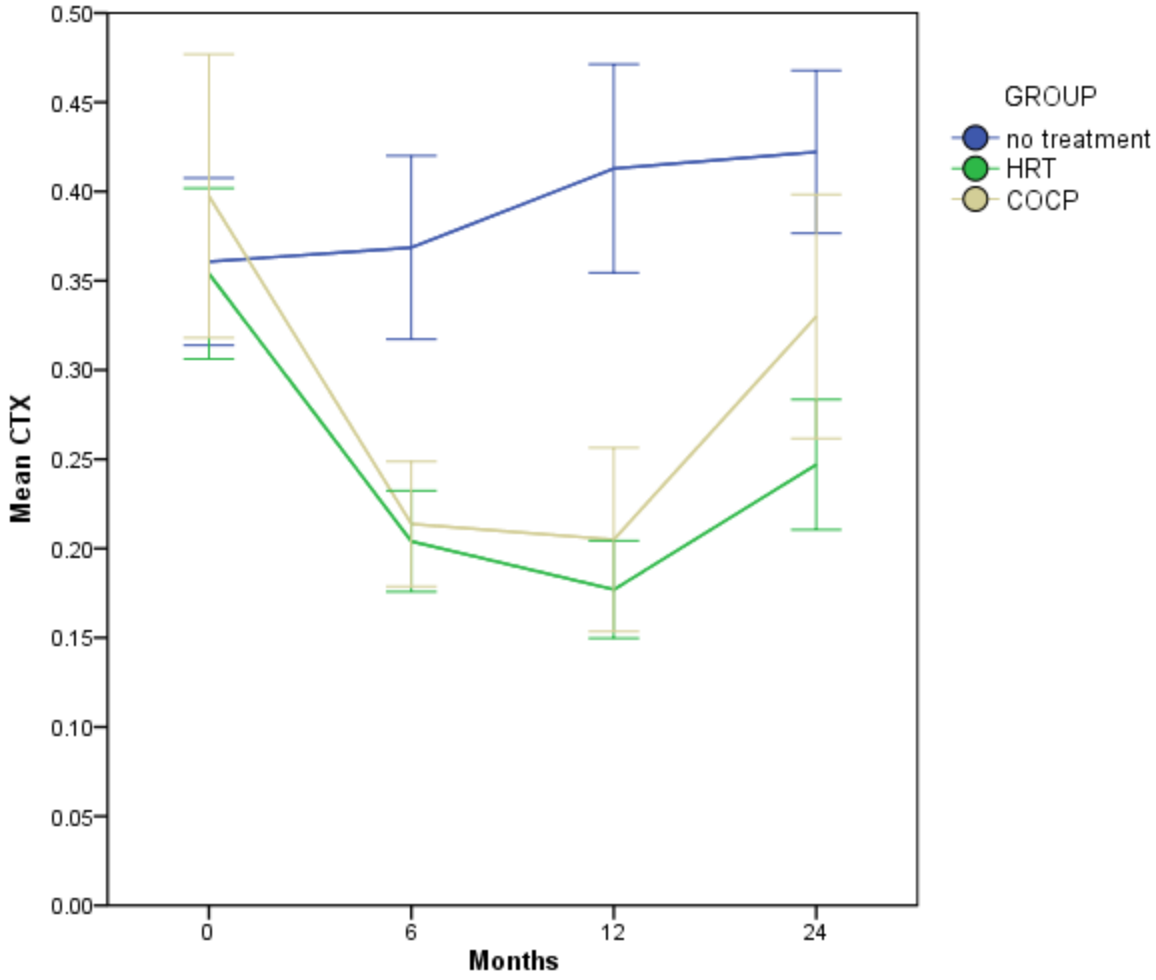


Figure 16 Changes in mean C-terminal cross-linked telopeptide (CTX) results (mcg/l) over 24 months in participants with complete data collection. Data shown as mean +/- 1 standard error.

4.12.2.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean CTX	95% confidence interval of the difference	p value
6	0.03	-0.04 to 0.10	0.367
12	0.02	-0.06 to 0.09	0.703
24	0.05	-0.08 to 0.17	0.451

Table 28 Comparison between HRT and COCP C-terminal cross-linked telopeptide (CTX) results (mcg/l). Linear regression analysis was used to adjust for baseline score

4.12.2.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean CTX	95% confidence interval of the difference	p value
6	-0.18	-0.29 to -0.07	0.003
12	-0.20	-0.30 to -0.11	<0.001
24	-0.16	-0.25 to -0.07	0.001

Table 29 Comparison between HRT and COCP C-terminal cross-linked telopeptide (CTX) results (mcg/l). Linear regression analysis was used to adjust for baseline score

4.12.2.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean CTX	95% confidence interval of the difference	p value
6	-0.07	-0.13 to -0.02	0.006
12	-0.10	-0.16 to -0.04	0.001
24	-0.06	-0.12 to 0.00	0.046

Table 30 Comparison between HRT and COCP C-terminal cross-linked telopeptide (CTX) results (mcg/l). Linear regression analysis was used to adjust for baseline score

4.13 Markers of ovarian function

4.13.1 Anti-Mullerian Hormone and Inhibin B

Anti-Mullerian hormone (AMH) was undetectable (less than 1.1 pmol/l) in all but one sample analysed. The exception was a level of 3.0pmol/l at baseline in a participant in the no treatment group. She withdrew following the 6 month visit, at which time her AMH was less than 1.1pmol/l. Her inhibin B levels were undetectable (<4.8 pg/ml) at baseline but increased to 255 at 6 months. Unfortunately the baseline samples for the participant in the no treatment group who became pregnant a few months into the trial were lost by the laboratory.

Inhibin B was detectable in 17 participants on at least one occasion. 11 participants had a detectable level on more than one occasion. This may represent a group of women with intermittent ovarian function, although in every case the AMH taken at the same time was undetectable. The split between the groups was roughly proportionate to the numbers in each group.

group	study number	Baseline	6 months	12 months	24 months
No treatment	6	14.3	undetectable	withdrew	withdrew
	17	17.1	undetectable	undetectable	undetectable
	22	31.9	16.5	27.1	undetectable
	26	undetectable	255.0	withdrew	withdrew
	27	undetectable	undetectable	22.6	undetectable
	28	undetectable	21.1	80.9	undetectable
	34	40.7	18.1	undetectable	undetectable
	36	5.4	no sample	undetectable	24.2
	52	25.9	7.7	13.8	6.7
HRT	32	10.5	48.9	undetectable	undetectable
	45	45.2	undetectable	undetectable	undetectable
	54	53.3	undetectable	undetectable	undetectable
	55	undetectable	23.6	undetectable	undetectable
COCP	29	24.1	withdrew	withdrew	withdrew
	41	undetectable	55.2	no sample	undetectable
	49	undetectable	51.4	64.2	undetectable
	58	6.2	undetectable	undetectable	withdrew

Table Table 31 Inhibin B over 24 months in participants with at least one detectable level (pg/ml)

4.13.2 Antral follicle count and ovarian volume

Due to unforeseeable circumstances the ultrasound machine became unavailable at the end of January 2010. The machine that replaced it was inadequate for accurately locating the ovaries and measuring the antral follicle count. This meant that baseline data was collected only on the first 29 participants. Six month data was collected from only 5 participants and therefore has

not been analysed. Baseline data is presented below. As expected in this population, data is skewed to the left and is therefore presented as median (25, 75%).

	Median (25, 75%)
Antral follicle count	5.0 (3.0, 7.8)
Ovarian volume (cm^3)	2.5 (1.3, 5.1)
Endometrial thickness (mm)	2.8 (2.0, 4.8)

Table 32 Baseline ultrasound data

Lipid profile

Several of the lipid profile parameters show trends in favour of HRT (for example, a reduction in low density lipoprotein in the HRT group, and a reduction in high density lipoprotein in the COCP group), but on formal statistical analysis few significant differences were found. A larger sample size would be needed to confirm a true difference.

4.13.3 Low density lipoprotein

Figs 17 and 18 appear to show a reduction in low density lipoprotein (LDL) in the HRT group at 6 months which is then maintained over the course of the trial, whereas the levels in the COCP and no treatment groups remain relatively constant. However, statistical analysis did not reveal any significant differences between the HRT and COCP groups.

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline LDL	15	3.13 (0.56)	15	2.76 (0.78)	29	2.65 (0.84)
Change from baseline to 6 months	12	-0.28 (0.59)	12	-0.02 (0.43)	23	+0.07 (0.53)
Change from baseline to 12 months	13	-0.22 (0.56)	10	+0.05 (0.26)	21	-0.20 (0.39)
Change from baseline to 24 months	12	-0.18 (0.57)	9	+0.01 (0.74)	15	+0.08 (0.54)

Table 33 Low density lipoprotein (LDL) results (mmol/l)

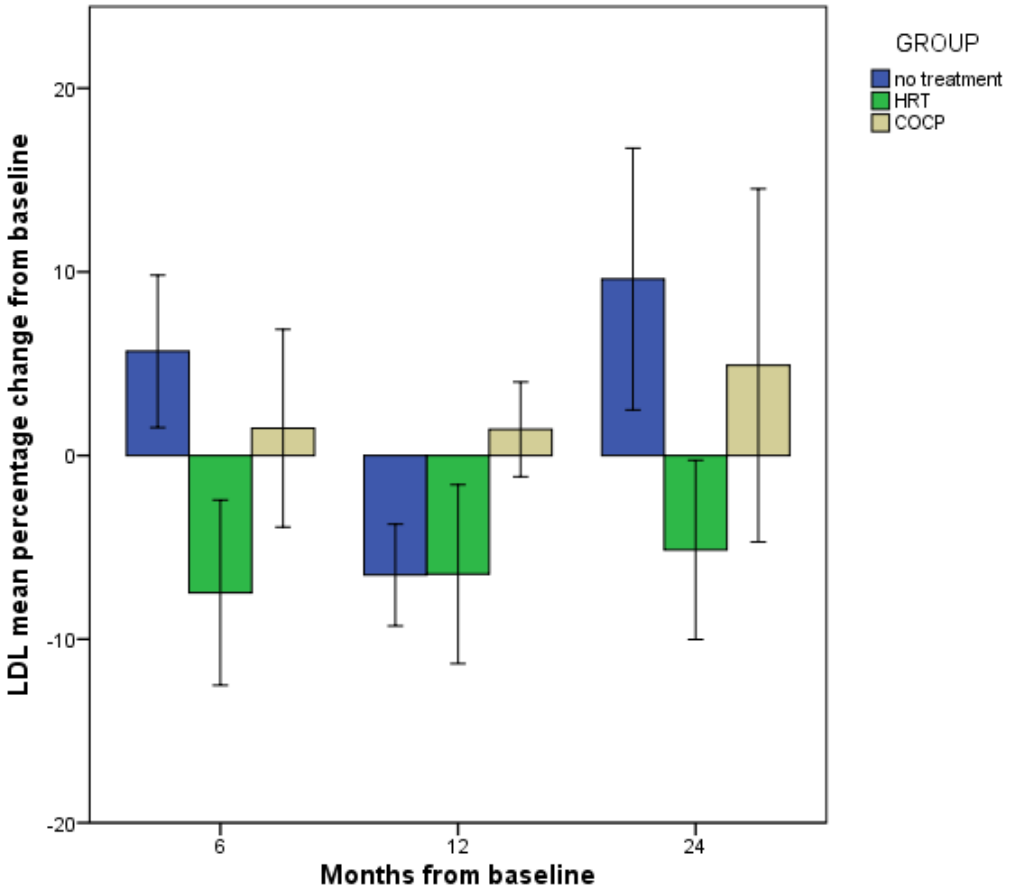


Figure 17 Percent changes from baseline in mean low density lipoprotein (LDL) over 24 months showing all available data at each time-point. Data shown as mean +/- 1 standard error.

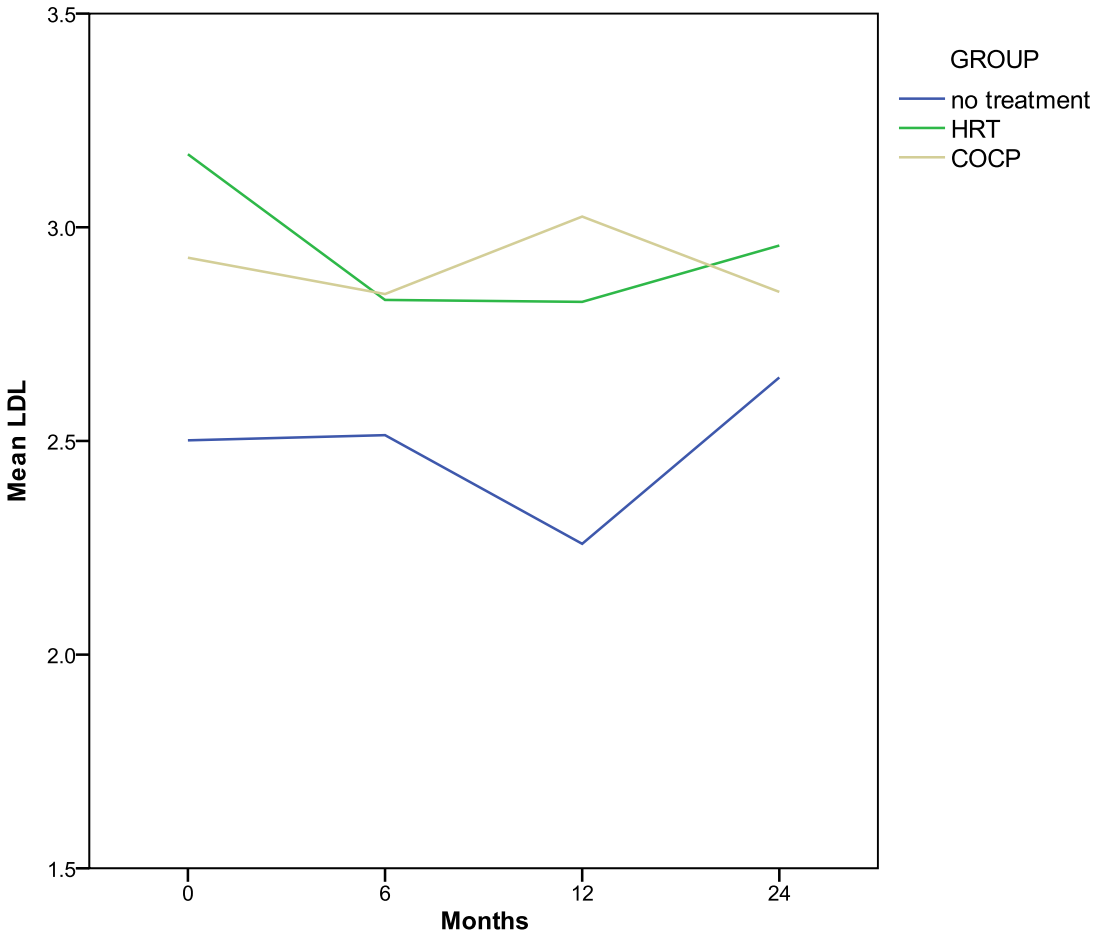


Figure 18 Changes in mean low density lipoprotein (LDL) (mmol/l) over 24 months in participants with complete data collection

4.13.3.1 Comparison between HRT and COCP

Months	COCP minus HRT mean LDL	95% confidence interval of the difference	p value
6	0.08	-0.30 to 0.47	0.657
12	0.24	-0.19 to 0.67	0.252
24	0.05	-0.51 to 0.62	0.843

Table 34 Comparison between HRT and COCP low density lipoprotein (LDL) results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.3.2 Comparison between HRT and no treatment

Months	HRT minus no treatment mean LDL	95% confidence interval of the difference	p value
6	-0.17	-0.53 to 0.18	0.329
12	0.16	-0.18 to 0.49	0.341
24	0.02	-0.40 to 0.43	0.928

Table 35 Comparison between HRT and no treatment low density lipoprotein (LDL) results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.3.3 Comparison between COCP and no treatment

Months	COCP minus no treatment mean LDL	95% confidence interval of the difference	p value
6	-0.03	-0.19 to 0.12	0.658
12	0.14	0.01 to 0.28	0.037
24	0.06	-0.14 to 0.26	0.532

Table 36 Comparison between COCP and no treatment low density lipoprotein (LDL) results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.4 High density lipoprotein

A reduction in high density lipoprotein (HDL) was seen in the COCP group from 6 months, whilst levels in the HRT and no treatment groups remained relatively stable. The differences between the COCP and HRT groups at 6 and 12 months were significant (p 0.001 and 0.030), but at 24 months the significance level was 0.091. Comparison between the HRT and no treatment groups did not reveal any statistically significant differences. There were highly significant differences between the COCP and no treatment groups at 6 and 12 months, and the difference remained significant at 24 months (p 0.049).

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline HDL	15	1.81 (0.44)	15	1.90 (0.50)	29	1.97 (0.52)
Change from baseline to 6 months	12	-0.06 (0.22)	12	-0.32 (0.22)	23	+0.03 (0.33)
Change from baseline to 12 months	13	-0.03 (0.28)	10	-0.28 (0.22)	21	+0.15 (0.26)
Change from baseline to 24 months	12	+0.04 (0.27)	9	-0.14 (0.25)	15	+0.06 (0.28)

Table 37 High density lipoprotein (HDL) results (mmol/l)

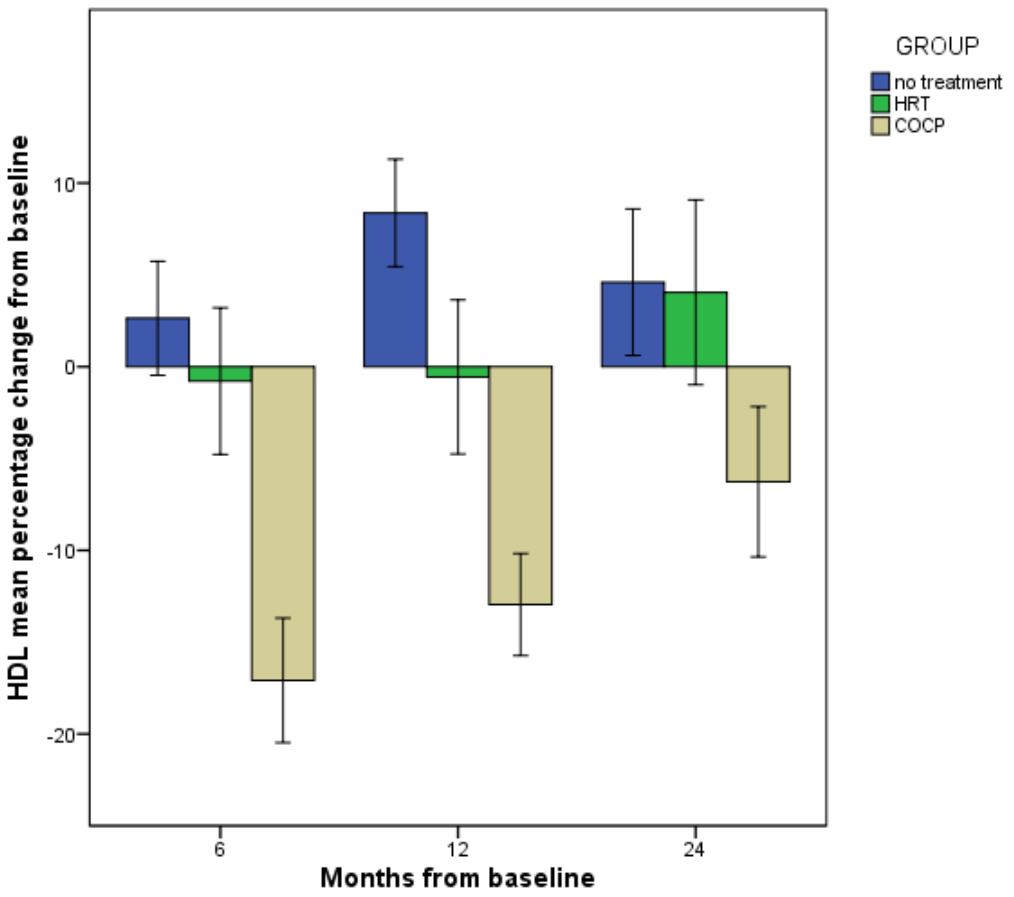


Figure 19 Percent changes from baseline in mean high density lipoprotein (HDL) over 24 months showing all available data at each time-point. Data shown as mean +/- 1 standard error.

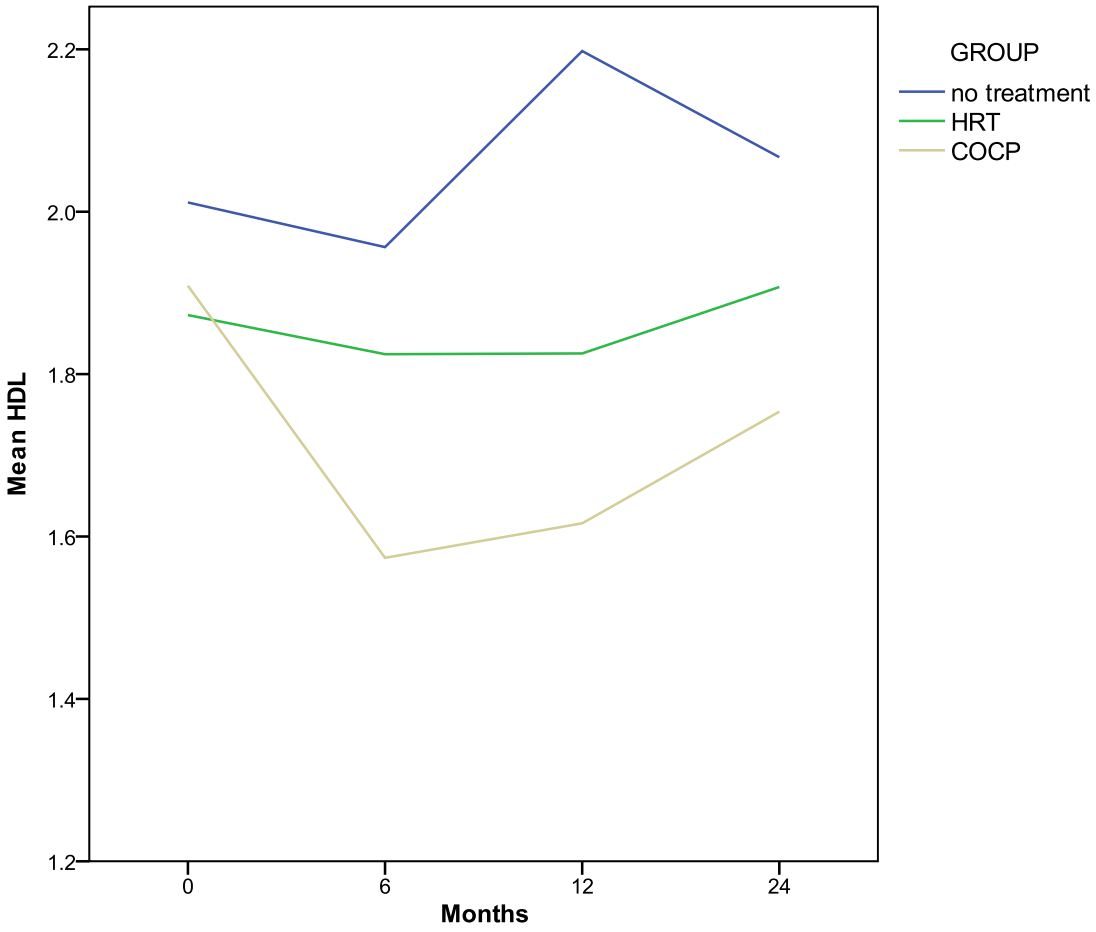


Figure 20 Changes in mean high density lipoprotein (HDL) (mmol/l) over 24 months in participants with complete data collection

4.13.4.1 Comparison between HRT and COCP

Months	COCP minus HRT mean HDL	95% confidence interval of the difference	p value
6	-0.27	-0.42 to -0.13	0.001
12	-0.22	-0.41 to -0.02	0.030
24	-0.19	-0.41 to 0.03	0.091

Table 38 Comparison between HRT and COCP high density lipoprotein (HDL) results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.4.2 Comparison between HRT and no treatment

Months	HRT minus no treatment mean HDL	95% confidence interval of the difference	p value
6	-0.11	-0.31 to 0.09	0.252
12	-0.19	-0.39 to 0.00	0.054
24	-0.03	-0.24 to 0.18	0.765

Table 39 Comparison between HRT and no treatment high density lipoprotein (HDL) results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.4.3 Comparison between COCP and no treatment

Months	COCP minus no treatment mean HDL	95% confidence interval of the difference	p value
6	-0.19	-0.29 to -0.09	0.001
12	-0.21	-0.31 to -0.12	<0.001
24	-0.11	-0.22 to -0.00	0.049

Table 40 Comparison between COCP and no treatment high density lipoprotein (HDL) results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.5 Total cholesterol

Figs 21 and 22 show a similar reduction in total cholesterol in both the HRT and COCP groups from 6 months, whereas in the no treatment group levels remain constant. However, there were no consistently significant differences between the groups.

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline total cholesterol	15	5.35 (0.83)	15	5.05 (0.97)	29	5.02 (0.85)
Change from baseline to 6 months	12	-0.35 (0.65)	12	-0.35 (0.41)	23	+0.05 (0.59)
Change from baseline to 12 months	13	-0.31 (0.65)	10	-0.25 (0.34)	21	-0.12 (0.51)
Change from baseline to 24 months	12	-0.17 (0.59)	9	-0.14 (0.81)	15	0.07 (0.61)

Table 41 Total cholesterol results (mmol/l)

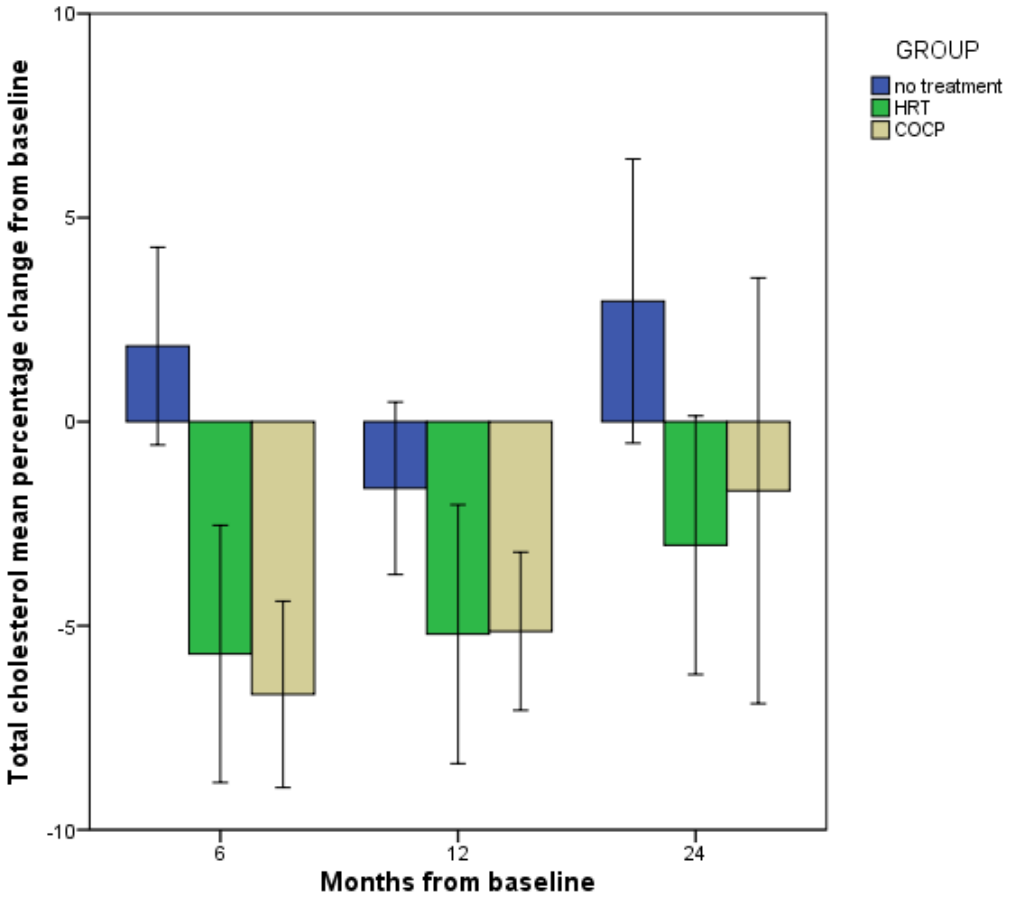


Figure 21 Percent changes from baseline in total cholesterol over 24 months showing all available data at each time-point. Data shown as mean +/- 1 standard error.

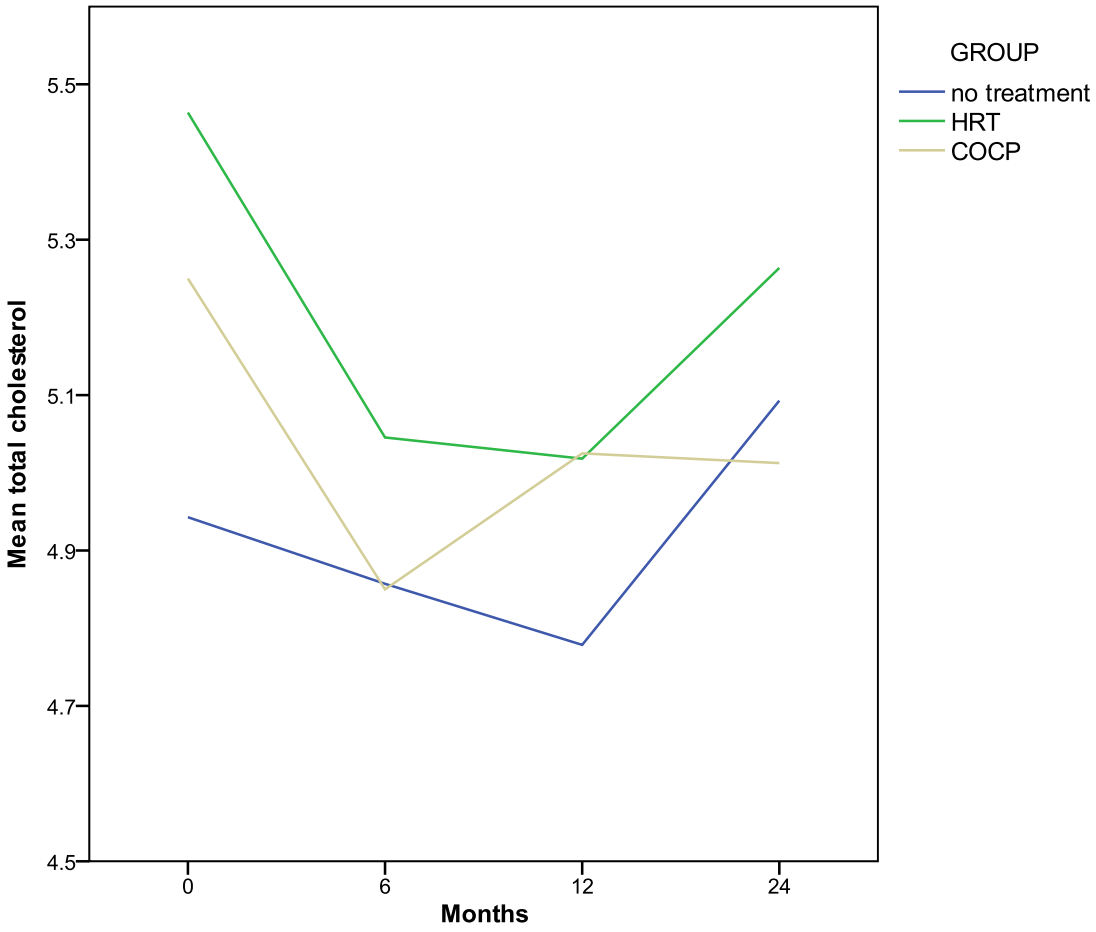


Figure 22 Changes in mean total cholesterol (mmol/l) over 24 months in participants with complete data collection

4.13.5.1 Comparison between HRT and COCP

Months	COCP minus HRT mean total cholesterol	95% confidence interval of the difference	p value
6	-0.14	-0.56 to 0.27	0.479
12	0.00	-0.47 to 0.48	0.993
24	-0.05	-0.70 to 0.60	0.874

Table 42 Comparison between HRT and COCP total cholesterol results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.5.2 Comparison between HRT and no treatment

Months	HRT minus no treatment mean total cholesterol	95% confidence interval of the difference	p value
6	-0.28	-0.68 to 0.11	0.156
12	0.02	-0.35 to 0.38	0.933
24	-0.07	-0.54 to 0.40	0.765

Table 43 Comparison between HRT and no treatment total cholesterol results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.5.3 Comparison between COCP and no treatment

Months	COCP minus no treatment mean total cholesterol	95% confidence interval of the difference	p value
6	-0.22	-0.39 to -0.04	0.019
12	-0.04	-0.21 to 0.13	0.611
24	-0.05	-0.30 to 0.20	0.667

Table 44 Comparison between COCP and no treatment total cholesterol results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.6 Triglycerides

Figs 23 and 24 indicate that triglycerides decreased in all groups, but most in the no treatment group. There were no significant differences between the groups.

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline triglycerides	15	0.92 (0.20)	15	0.89 (0.31)	29	0.91 (0.45)
Change from baseline to 6 months	12	-0.03 (0.42)	12	-0.03 (0.38)	23	-0.10 (0.31)
Change from baseline to 12 months	13	-0.11 (0.21)	10	-0.06 (0.46)	21	-0.15 (0.31)
Change from baseline to 24 months	12	-0.07 (0.40)	9	-0.03 (0.53)	15	-0.14 (0.34)

Table 45 Triglyceride results (mmol/l)

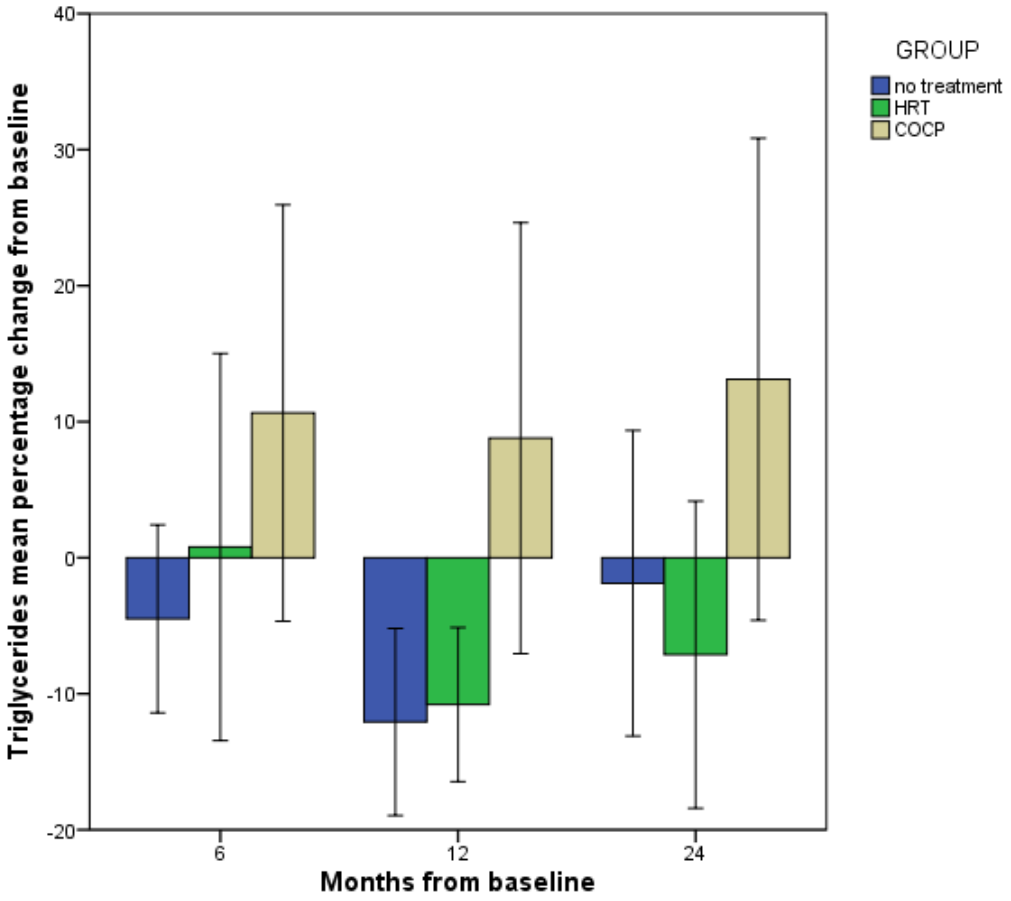


Figure 23 Percent changes from baseline in triglycerides over 24 months showing all available data at each time-point. Data shown as mean +/- 1 standard error

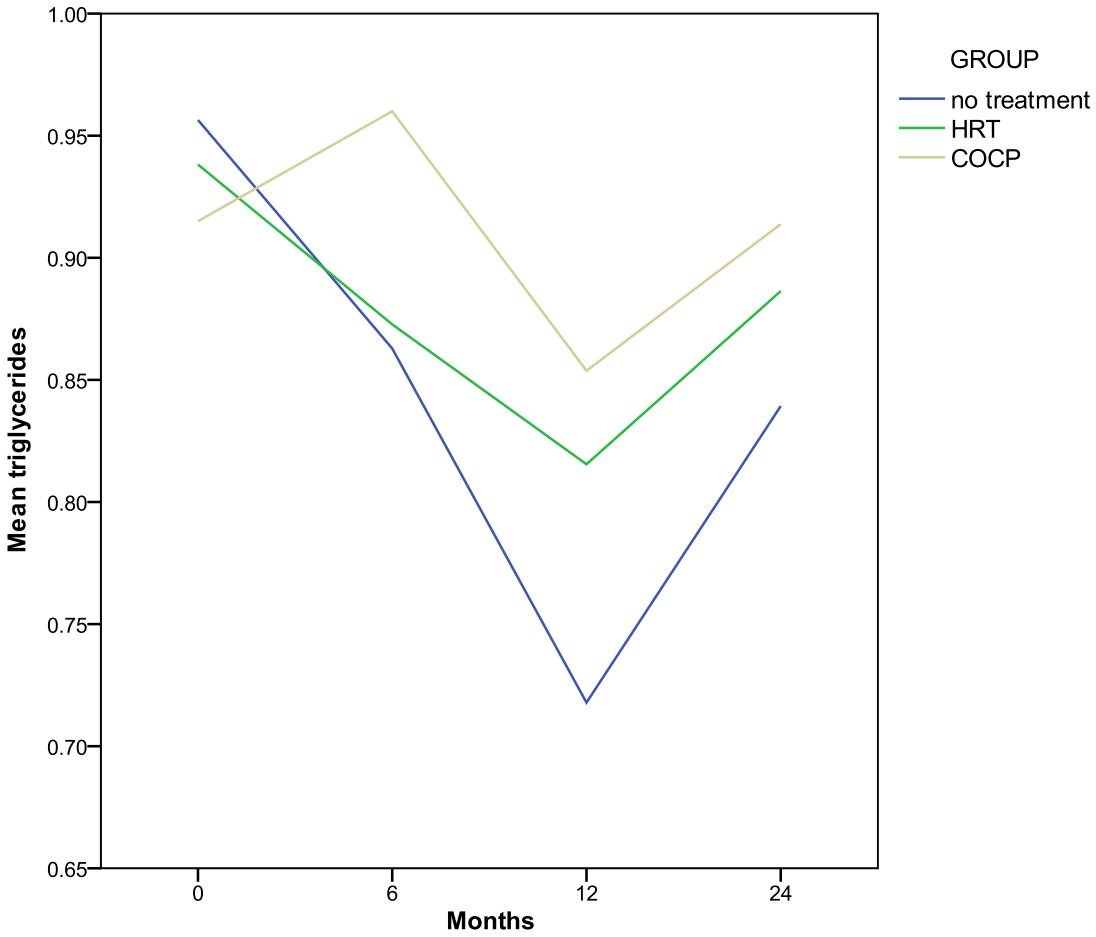


Figure 24 Changes in mean triglycerides (mmol/l) over 24 months in participants with complete data collection

4.13.6.1 Comparison between HRT and COCP

Months	COCP minus HRT mean triglycerides	95% confidence interval of the difference	p value
6	-0.00	-0.25 to 0.24	0.987
12	0.04	-0.14 to 0.22	0.625
24	0.03	-0.32 to 0.39	0.842

Table 46 Comparison between HRT and COCP triglycerides results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.6.2 Comparison between HRT and no treatment

Months	HRT minus no treatment mean triglycerides	95% confidence interval of the difference	p value
6	0.06	-0.17 to 0.29	0.593
12	0.03	-0.15 to 0.20	0.754
24	0.05	-0.18 to 0.29	0.642

Table 47 Comparison between HRT and no treatment triglycerides results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.13.6.3 Comparison between COCP and no treatment

Months	COCP minus no treatment mean triglycerides	95% confidence interval of the difference	p value
6	0.03	-0.07 to 0.13	0.536
12	0.04	-0.08 to 0.15	0.523
24	0.04	-0.07 to 0.15	0.450

Table 48 Comparison between COCP and no treatment triglycerides results (mmol/l). Linear regression analysis was used to adjust for baseline score

4.14 High Sensitivity C-Reactive Protein

The high sensitivity C-reactive protein (hsCRP) results were not normally distributed and therefore data are shown as median (25, 75%). Attempts to transform the data into a normal distribution, including using logs, inverse, squaring and square roots were not successful in achieving a normal distribution. Following statistical advice, the non-parametric Mann Whitney-U test was used to compare changes between groups. One woman in the no treatment group had a very high hsCRP due to sarcoidosis and arthritis and her results were excluded prior to analysis. Problems with analysis in the laboratory meant that there were more missing data in the hsCRP results than elsewhere.

Fig 25 shows that the hsCRP increased at all time-points in the COCP group but remained relatively unchanged in the HRT and no treatment groups. The differences between the groups were significant at all time-points when comparing the COCP and no treatment groups, and at 24 months between the COCP and HRT groups. However, these results need to be interpreted with caution for two reasons. Firstly, the numbers are very small, especially at 24 months in the COCP group. Secondly, the COCP group had a lower hsCRP than the other groups at baseline and this was even more marked amongst women who completed the trial, as shown in fig 26. Various transformations of the data did not produce a normal distribution and therefore it was not possible to use linear regression for analysis. It is not possible to control for the baseline score with non-parametric tests.

	HRT		COCP		No treatment	
	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)
Baseline hsCRP	15	1.10 (0.40, 1.60)	13	0.40 (0.20, 0.70)	25	0.80 (0.45, 1.95)
Change from baseline to 6 months	11	+0.10 (-1.00, 0.60)	10	+0.50 (0.20, 0.50)	20	-0.20 (-0.63, 0.20)
Change from baseline to 12 months	10	-0.22 (-0.70, 2.03)	8	+0.60 (0.33, 0.95)	12	-0.05 (-0.48, 0.28)
Change from baseline to 24 months	11	-0.10 (-1.10, 0.60)	6	+0.75 (0.38, 4,4)	12	-0.05 (-0.28, 0.28)

Table 49 High sensitivity C-reactive protein (hsCRP) results (mg/l)

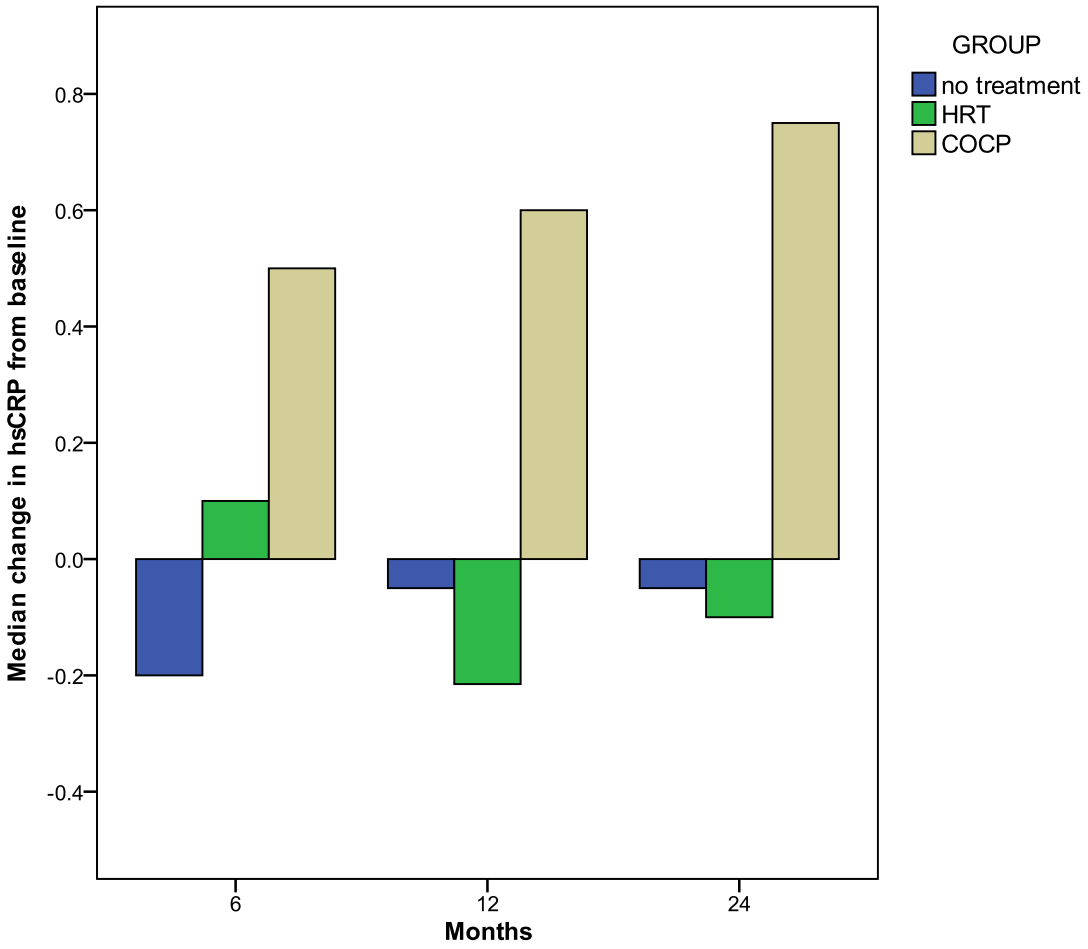


Figure 25 Changes from baseline in high sensitivity C-reactive protein (hsCRP) (mg/l) median score over 24 months showing all available data at each time-point

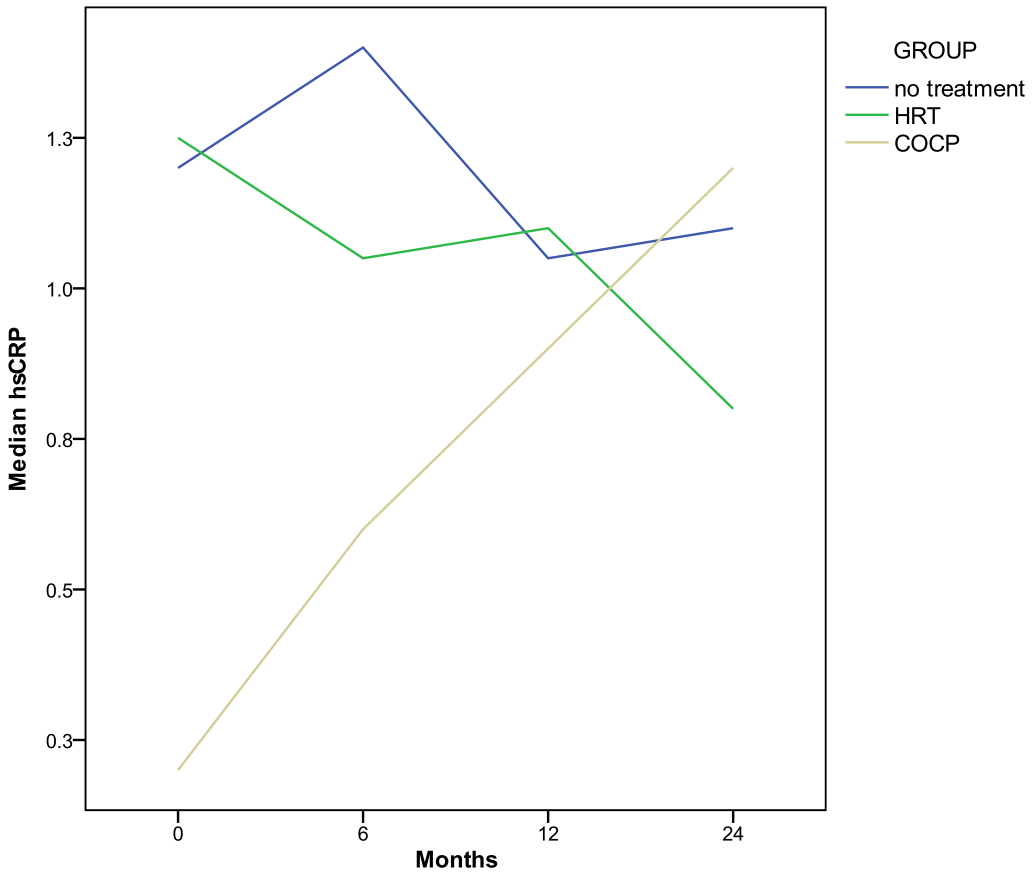


Figure 26 Changes from baseline in high sensitivity C-reactive protein median score (hsCRP) (mg/l) in participants with complete data collection

4.14.1 Comparison between groups

Mann-Whitney U tests showed a significant difference between the changes in score in the HRT and COCP groups at 24 months, but not at 6 or 12 months (p values 0.078, 0.307 and 0.035 at 6, 12 and 24 months respectively). There were no significant differences between the changes in the HRT and no treatment groups (p values 0.352, 0.869 and 0.622 at 6, 12 and 24 months). However, comparison between the COCP and no treatment groups revealed significant differences at all time points (p values <0.001, 0.018 and 0.009 at 6, 12 and 24 months).

4.15 Modified Greene Climacteric Scale

The Modified Greene Climacteric Scale (MGCS) results between HRT and COCP groups are difficult to interpret because although the scores were similar at baseline (as expected following randomisation), the women who completed the trial in the COCP group had higher baseline scores than women who completed the trial in the HRT group (see figure 28 below). The higher drop-out rate in the COCP group was due to loss of follow up, and there were just two reported withdrawals in the treatment groups due to menopausal symptoms - one each in the HRT and COCP groups. However, it is possible that loss of follow up in either group could have been due to problems with symptom control. Due to differences in baseline scores between participants who completed the trial and those who dropped out, graphs illustrating both all data collected and those with complete data collection are shown here and for the other questionnaires.

At 3 and 6 months there are good amounts of data from both the COCP and HRT groups. The results at 3 months include data from 14/15 participants in the HRT group and 13/15 in the COCP group. At this time point, the reductions in the total, psychological, anxiety, somatic and vasomotor scores are statistically significantly greater in the HRT group. However, at 6 months, with data available in 13/15 in the HRT group and 12/15 in the COCP group, there are no significant differences in scores. It appears that the COCP takes longer to control symptoms but by 6 months the results are comparable. This is supported by the graphs that present data from women who completed the trial and attended all visits (fig 28, 30, 32, 24 and 36); in these women a sharp reduction in score is seen in the HRT group at 3 months and the score then stabilises, whereas in the COCP group the reduction in score continues from 3 to 6 months. This is seen in each domain except vasomotor and sexual dysfunctions scores. In the vasomotor score, HRT and the COCP follow a similar pattern of rapid symptom reduction at 3 months then stabilisation of score. Scores are similar in each domain except somatic at 12 months. However, by 24 months there are significant differences between the HRT and COCP groups in all scores except sexual dysfunction and vasomotor scores. All differences are in favour of HRT. There are no significant differences between the HRT and COCP groups at any time-point in the sexual dysfunction score.

The regression analyses comparing the groups control for baseline score, and therefore estimates of differences between the groups are markedly different from the changes from baseline shown in the tables.

In the no treatment group, there was a high drop-out rate and several of these were due to menopausal symptoms. Therefore the results from this group should be interpreted as what happens to women who choose to continue on no treatment. It should also be noted that the no treatment group had significantly lower scores at baseline. The no treatment group was not randomised and therefore comparisons with the treatment groups are less valuable than comparisons between the two treatment groups. However, they illustrate the large differences between the HRT and no treatment groups and the few differences between the COCP and no treatment groups. The comparison between HRT and no treatment showed highly significant

reductions in score in the HRT group at almost all time-points and in all domains except sexual function. In contrast, the COCP versus no treatment comparison did not show a significant difference in total, psychological, anxiety, depression or somatic score at any time-point. There was a trend towards reduction in vasomotor symptoms in the COCP group compared with no treatment, but this only reached a significance level of less than 0.05 at 6 and 12 months. This contrasts sharply with the results in the HRT versus no treatment vasomotor score comparisons.

4.15.1 Total score

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline MGCS total score	15	22.4 (17.0)	15	24.2 (13.7)	29	15.8 (11.6)
Change from baseline to 3 months	14	-12.0 (11.6)	13	-3.5 (9.1)	25	-1.3 (7.7)
Change from baseline to 6 months	13	-9.8 (12.2)	12	-9.1 (11.2)	23	-1.4 (6.9)
Change from baseline to 12 months	13	-11.2 (12.5)	11	-7.4 (13.7)	21	+0.8 (9.2)
Change from baseline to 18 months	11	-8.2 (11.3)	10	-6.5 (15.1)	15	+0.7 (6.6)
Change from baseline to 24 months	12	-11.8 (13.9)	9	-7.6 (10.7)	15	-1.9 (8.1)

Table 50 Modified Greene Climacteric Scale (MGCS) total score results

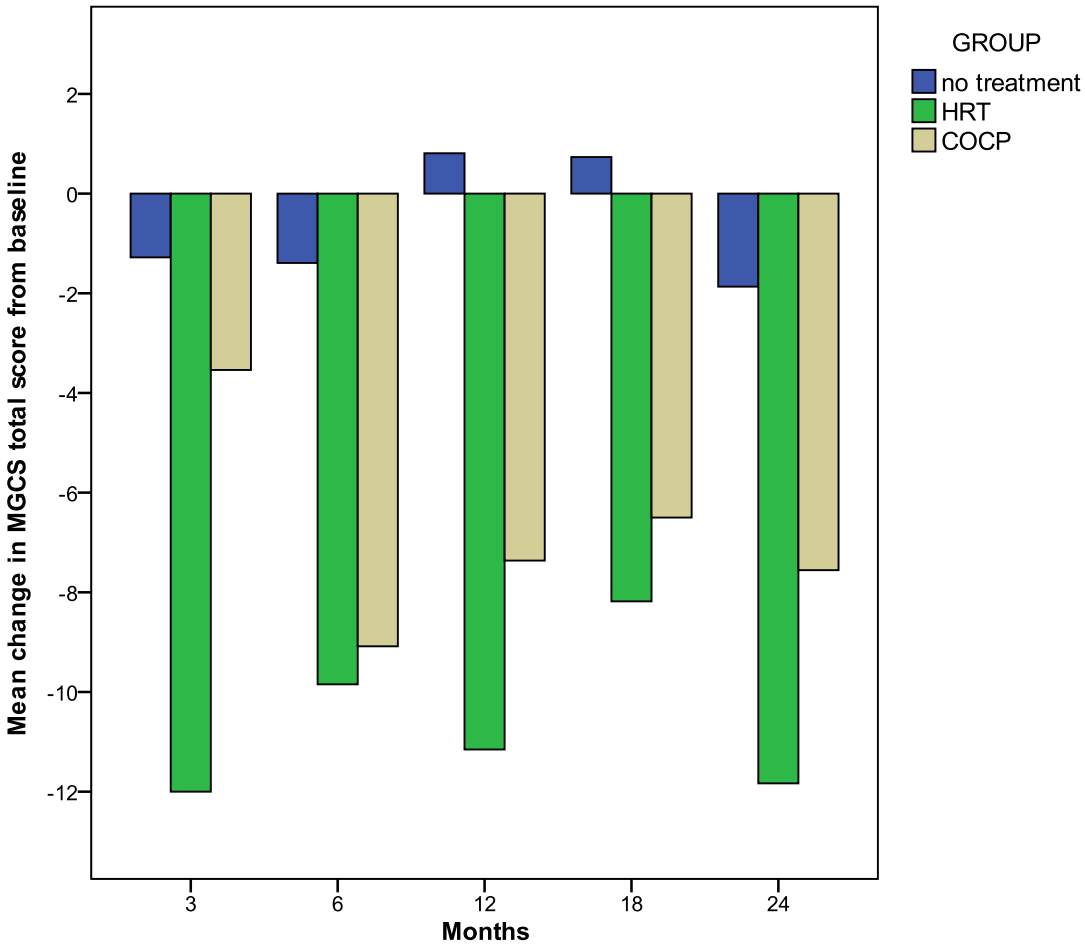


Figure 27 Changes from baseline in total Modified Greene Climacteric Scale (MGCS) mean score over 24 months showing all available data at each time-point

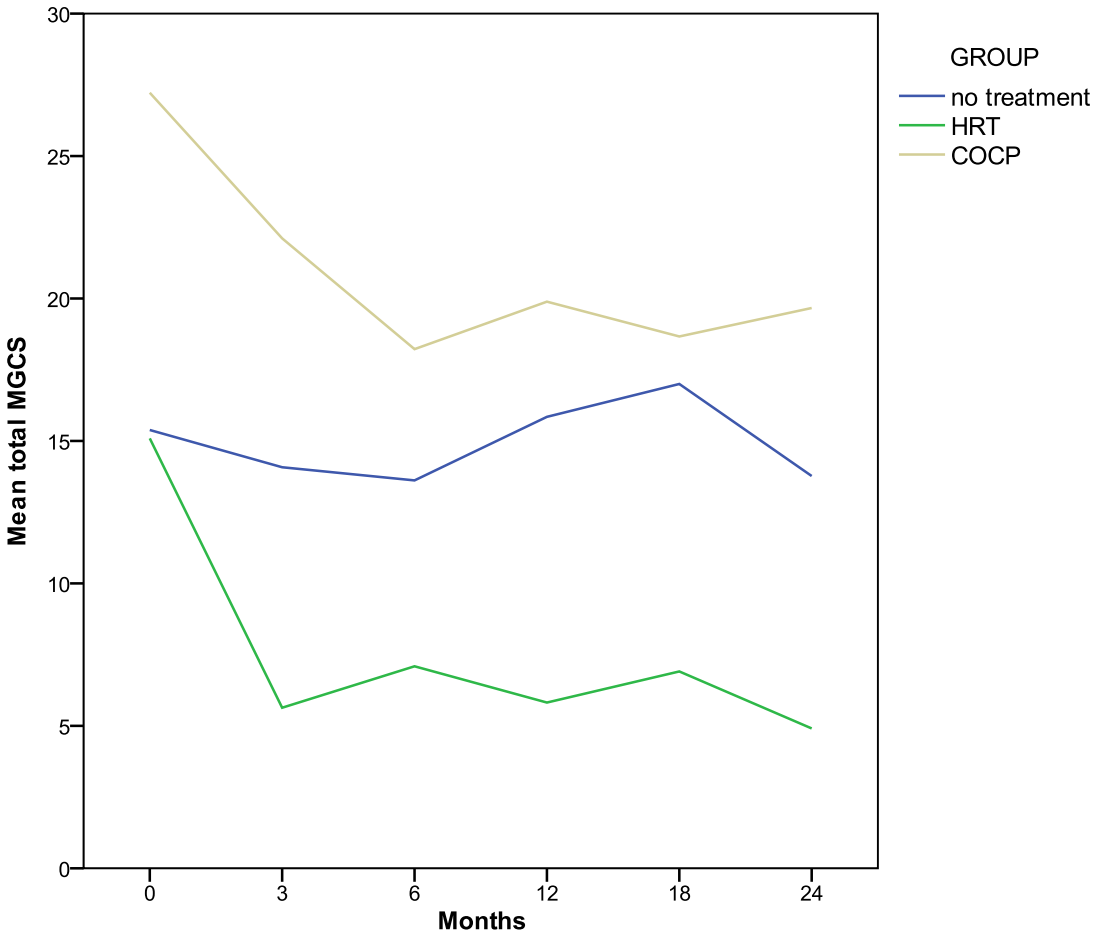


Figure 28 Changes in total Modified Greene Climacteric Scale (MGCS) over 24 months in participants with complete data collection

4.15.1.1 Comparison between HRT and COCP

Months	COCP minus HRT mean total MGCS results	95% confidence interval of the difference	p value
3	9.5	3.7 to 15.2	0.002
6	4.0	-2.4 to 10.3	0.205
12	7.5	-0.4 to 15.4	0.062
18	9.6	1.8 to 17.5	0.018
24	10.4	3.7 to 17.0	0.004

Table 51 Comparison between HRT and COCP Modified Greene Climacteric Scale (MGCS) total score results. Linear regression analysis was used to adjust for baseline score

4.15.1.2 Comparison between HRT and no treatment

Months	HRT minus no treatment mean total MGCS results	95% confidence interval of the difference	p value
3	-7.9	-11.8 to -4.0	<0.001
6	-5.7	-10.0 to -1.4	0.011
12	-8.6	-14.4 to -2.7	0.005
18	-9.1	-14.1 to -4.1	0.001
24	-8.6	-14.8 to -2.6	0.007

Table 52 Comparison between HRT and no treatment total Modified Greene Climacteric Scale (MGCS) score results. Linear regression analysis was used to adjust for baseline score.

4.15.1.3 Comparison between COCP and no treatment

Months	COCP minus no treatment mean total MGCS results	95% confidence interval of the difference	p value
3	0.4	-2.0 to 2.7	0.759
6	-1.4	-4.0 to 1.2	0.285
12	-1.2	-5.1 to 2.8	0.552
18	-1.1	-4.8 to 2.7	0.569
24	-0.6	-4.2 to 3.0	0.715

Table 53 Comparison between COCP and no treatment total Modified Greene Climacteric Scale (MGCS) score results. Linear regression analysis was used to adjust for baseline score.

4.15.2 Psychological score

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline MGCS psychological score	15	12.4 (10.3)	15	13.1 (8.4)	29	9.2 (6.8)
Change from baseline to 3 months	14	-6.2 (7.5)	13	-0.5 (4.9)	25	-1.2 (5.3)
Change from baseline to 6 months	13	-5.1 (7.3)	12	-3.7 (6.6)	23	-0.8 (4.7)
Change from baseline to 12 months	13	-5.7 (7.3)	11	-3.3 (6.6)	21	+1.3 (4.7)
Change from baseline to 18 months	11	-4.5 (7.3)	10	-3.2 (7.9)	15	+0.6 (5.1)
Change from baseline to 24 months	12	-6.2 (8.7)	9	-4.0 (6.6)	15	-0.7 (5.8)

Table 54 Modified Greene Climacteric Scale (MGCS) psychological score results

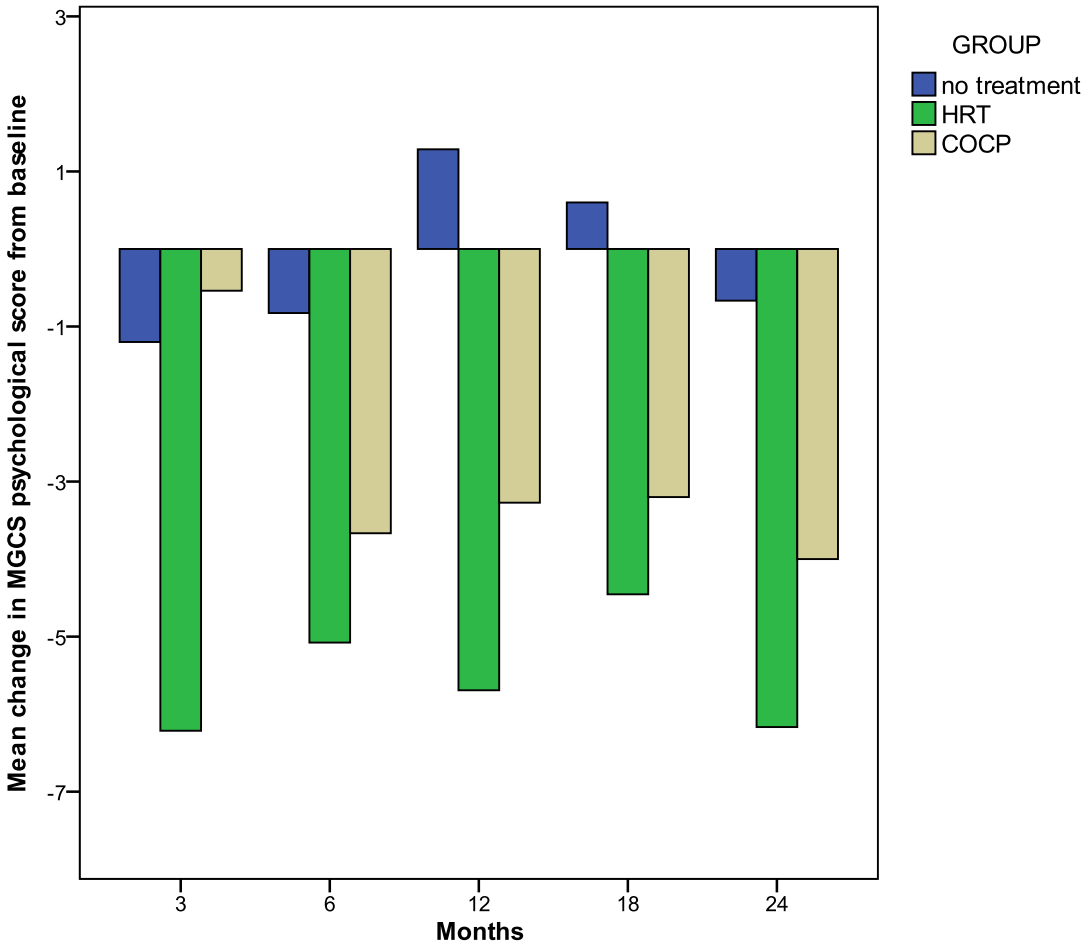


Figure 29 Changes from baseline in Modified Greene Climacteric Scale (MGCS) psychological mean score over 24 months showing all available data at each time-point

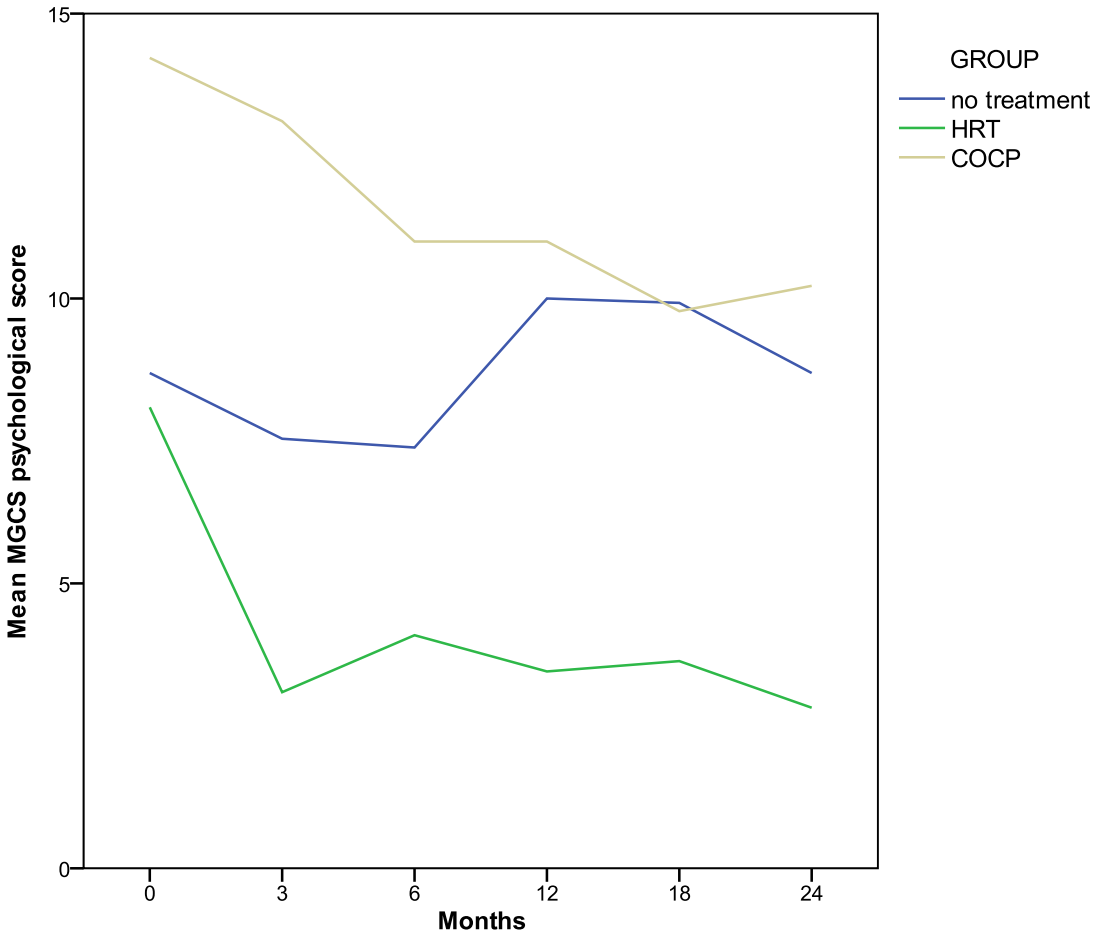


Figure 30 Changes in Modified Greene Climacteric Scale (MGCS) psychological score over 24 months in participants with complete data collection

4.15.2.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean MGCS psychological results	95% confidence interval of the difference	p value
3	5.9	2.4 to 9.3	0.002
6	3.0	-0.8 to 6.8	0.116
12	4.0	-0.4 to 8.4	0.071
18	5.1	0.9 to 9.4	0.021
24	5.2	1.1 to 9.3	0.015

Table 55 Comparison between HRT and COCP Modified Greene Climacteric Scale (MGCS) psychological score results. Linear regression analysis was used to adjust for baseline score.

Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean MGCS psychological results	95% confidence interval of the difference	p value
3	-3.6	-6.5 to -0.8	0.013
6	-2.9	-5.7 to -0.1	0.041
12	-5.5	-8.7 to -2.2	0.002
18	-5.7	-8.8 to -2.6	0.001
24	-5.1	-9.0 to -1.2	0.012

Table 56 Comparison between HRT and no treatment Modified Greene Climacteric Scale (MGCS) psychological score results. Linear regression analysis was used to adjust for baseline score.

4.15.2.2 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean MGCS psychological results	95% confidence interval of the difference	p value
3	1.0	-0.5 to 2.4	0.186
6	-0.3	-2.0 to 1.4	0.725
12	-1.2	-3.3 to 0.8	0.220
18	-0.7	-2.8 to 1.3	0.467
24	-0.5	-2.8 to 1.7	0.637

Table 57 Comparison between COCP and no treatment Modified Greene Climacteric Scale (MGCS) psychological score results. Linear regression analysis was used to adjust for baseline score.

4.15.3 Anxiety score

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline MGCS anxiety score	15	6.3 (5.3)	15	6.5 (4.3)	29	4.2 (3.3)
Change from baseline to 3 months	14	-3.7 (4.4)	13	0.0 (2.8)	25	0.0 (2.8)
Change from baseline to 6 months	13	-2.5 (3.6)	12	-1.8 (4.0)	23	-0.1 (2.5)
Change from baseline to 12 months	13	-3.1 (4.3)	11	-1.6 (4.4)	21	+1.0 (3.0)
Change from baseline to 18 months	11	-1.8 (3.8)	10	-0.9 (4.1)	15	+0.9 (2.5)
Change from baseline to 24 months	12	-3.2 (4.9)	9	-1.6 (3.4)	15	+0.1 (3.1)

Table 58 Modified Greene Climacteric Scale (MGCS) anxiety score results

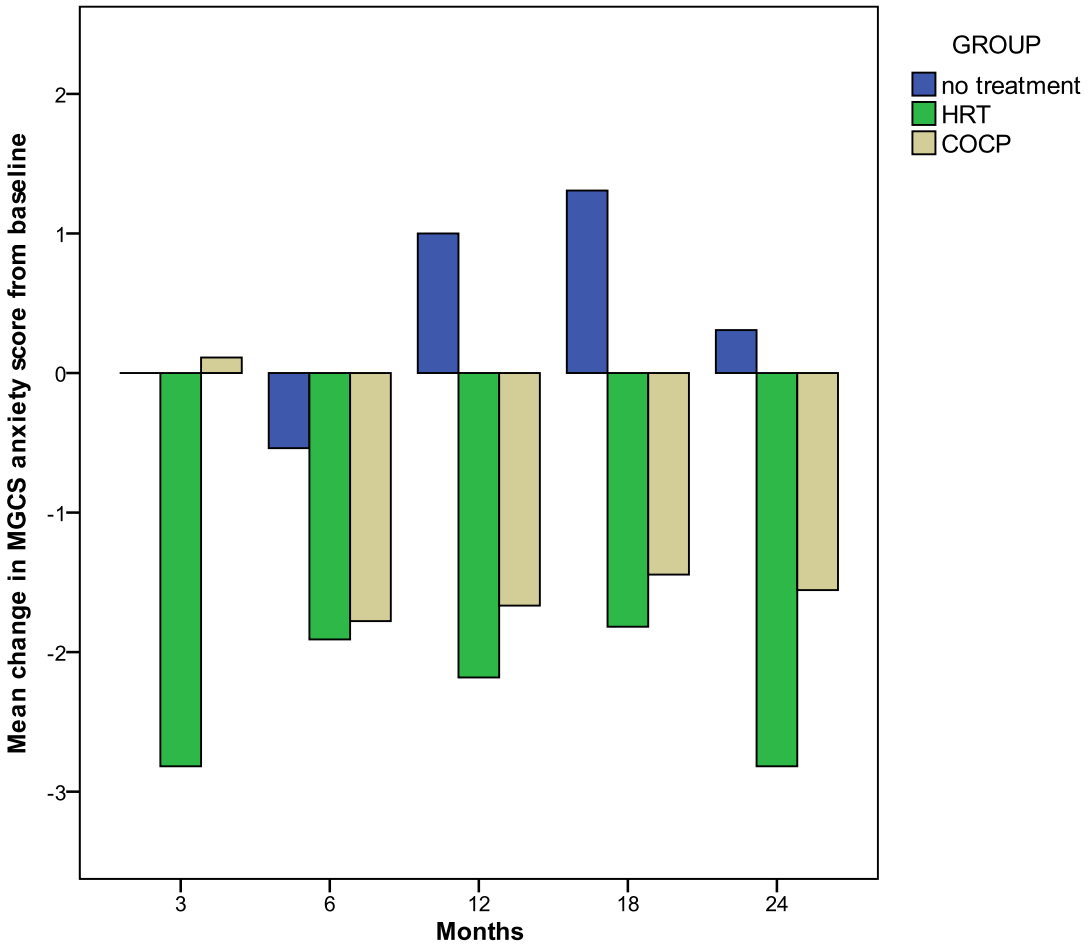


Figure 31 Changes from baseline in Modified Greene Climacteric Scale (MGCS) anxiety mean score over 24 months showing all available data at each time-point

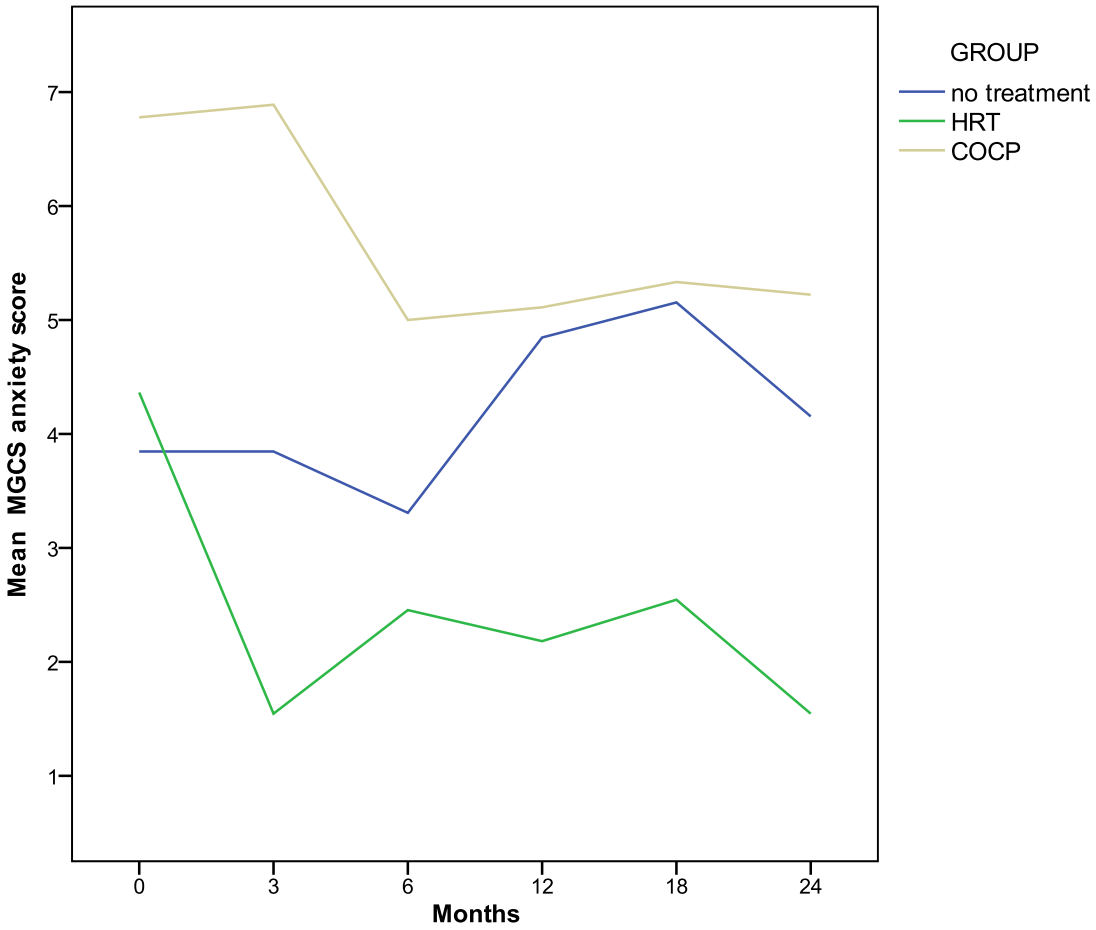


Figure 32 Changes in Modified Greene Climacteric Scale (MGCS) anxiety score over 24 months in participants with complete data collection

4.15.3.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean MGCS anxiety results	95% confidence interval of the difference	p value
3	3.8	2.0 to 5.5	<0.001
6	1.4	-0.6 to 3.5	0.167
12	2.1	-0.1 to 4.4	0.064
18	2.3	0.0 to 4.6	0.052
24	2.7	0.2 to 5.2	0.038

Table 59 Comparison between HRT and COCP Modified Greene Climacteric Scale (MGCS) anxiety score results. Linear regression analysis was used to adjust for baseline score.

4.15.3.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean MGCS anxiety results	95% confidence interval of the difference	p value
3	-2.5	-4.0 to -1.0	0.002
6	-1.4	-2.8 to 0.0	0.056
12	-2.7	-4.7 to -0.6	0.011
18	-2.6	-4.5 to -0.7	0.009
24	-2.6	-4.9 to -0.3	0.031

Table 60 Comparison between HRT and no treatment Modified Greene Climacteric Scale (MGCS) anxiety score results. Linear regression analysis was used to adjust for baseline score.

4.15.3.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean MGCS anxiety results	95% confidence interval of the difference	p value
3	0.5	-0.4 to 1.3	0.273
6	-0.0	-1.0 to 0.9	0.923
12	-0.5	-1.8 to 0.8	0.465
18	-0.4	-1.5 to 0.8	0.540
24	-0.3	-1.7 to 1.1	0.630

Table 61 Comparison between COCP and no treatment Modified Greene Climacteric Scale (MGCS) anxiety score results. Linear regression analysis was used to adjust for baseline score.

4.15.4 Depression score

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline MGCS depression score	15	6.2 (5.1)	15	6.5 (4.7)	29	5.0 (3.9)
Change from baseline to 3 months	14	-2.6 (3.4)	13	-0.5 (3.1)	25	-1.1 (3.0)
Change from baseline to 6 months	13	-2.6 (3.9)	12	-1.8 (2.9)	23	-0.7 (2.6)
Change from baseline to 12 months	13	-2.7 (3.4)	11	-1.6 (2.7)	21	+0.3 (2.4)
Change from baseline to 18 months	11	-2.7 (3.7)	10	-2.3 (4.3)	15	+0.7 (4.0)
Change from baseline to 24 months	12	-3.0 (4.0)	9	-2.4 (3.9)	15	-0.9 (3.3)

Table 62 Modified Greene Climacteric Scale (MGCS) depression score results

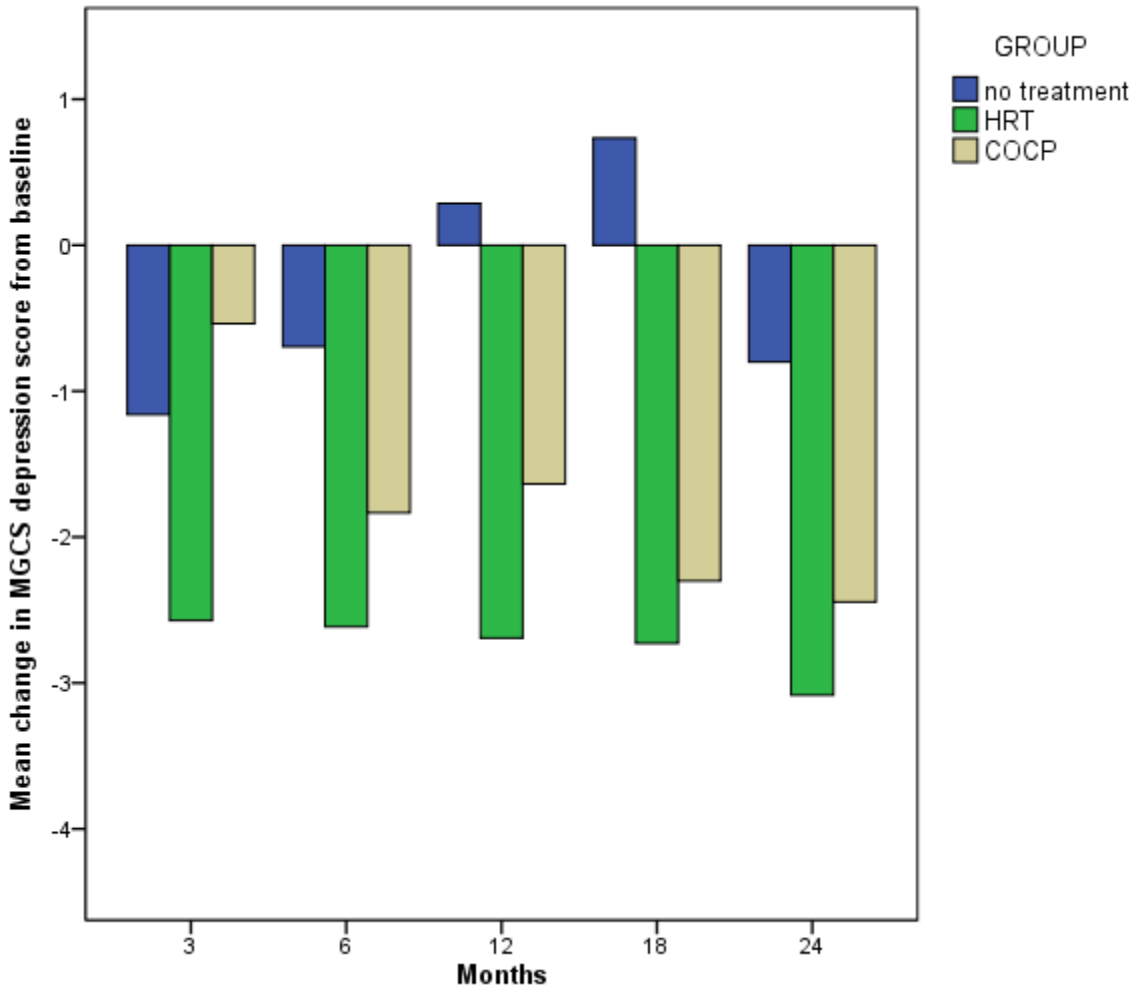


Figure 33 Changes from baseline in Modified Greene Climacteric Scale (MGCS) depression mean score over 24 months showing all available data at each time-point

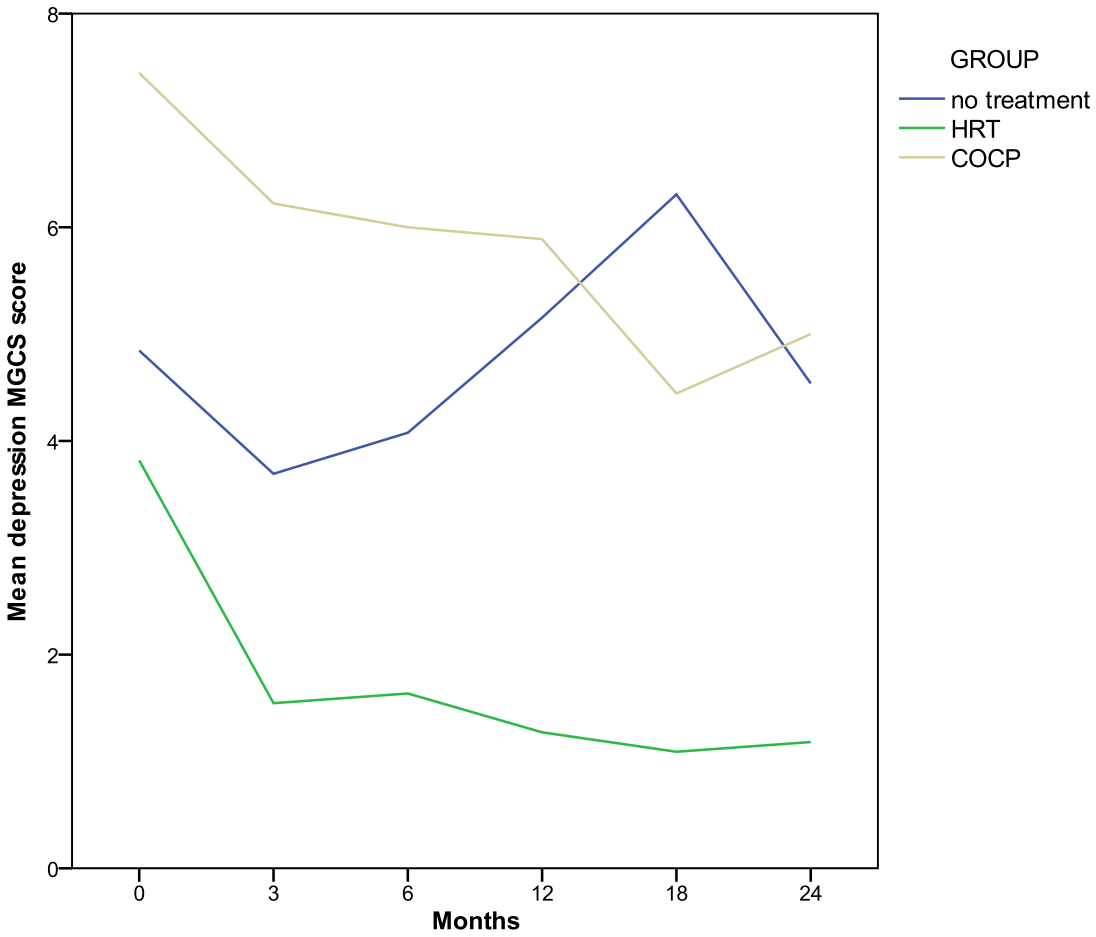


Figure 34 Changes in Modified Greene Climacteric Scale (MGCS) depression score over 24 months in participants with complete data collection

4.15.4.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean MGCS depression results	95% confidence interval of the difference	p value
3	2.1	-0.06 to 4.2	0.056
6	1.6	-0.5 to 3.8	0.133
12	1.7	-0.7 to 4.1	0.149
18	2.8	0.4 to 5.2	0.027
24	2.6	0.5 to 4.7	0.017

Table 63 Comparison between HRT and COCP Modified Greene Climacteric Scale (MGCS) depression score results. Linear regression analysis was used to adjust for baseline score.

4.15.4.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean MGCS depression results	95% confidence interval of the difference	p value
3	-1.1	-2.7 to 0.5	0.177
6	-1.6	-3.2 to 0.1	0.064
12	-2.6	-4.3 to -1.0	0.002
18	-3.9	-7.0 to -0.8	0.017
24	-2.7	-4.6 to -0.8	0.008

Table 64 Comparison between HRT and no treatment Modified Greene Climacteric Scale (MGCS) depression score results. Linear regression analysis was used to adjust for baseline score.

4.15.4.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean MGCS depression results	95% confidence interval of the difference	p value
3	0.5	-0.3 to 1.3	0.240
6	-0.1	-1.0 to 1.7	0.755
12	-0.5	-1.4 to 0.3	0.216
18	-1.2	-3.0 to 0.5	0.157
24	-0.2	-1.4 to 0.9	0.667

Table 65 Comparison between COCP and no treatment Modified Greene Climacteric Scale (MGCS) depression score results. Linear regression analysis was used to adjust for baseline score.

4.15.5 Somatic score

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline MGCS somatic score	15	6.3 (5.5)	15	6.0 (4.4)	29	4.0 (4.2)
Change from baseline to 3 months	14	-3.1 (3.2)	13	-0.4 (3.4)	25	-0.4 (2.4)
Change from baseline to 6 months	13	-2.7 (3.1)	12	-2.2 (3.8)	23	-1.0 (3.2)
Change from baseline to 12 months	13	-2.9 (3.1)	11	-1.5 (4.5)	21	-0.3 (2.6)
Change from baseline to 18 months	11	-1.7 (2.0)	10	-0.9 (5.1)	15	-0.8 (2.4)
Change from baseline to 24 months	12	-3.3 (3.2)	9	-1.0 (2.3)	15	-0.9 (2.7)

Table 66 Modified Greene Climacteric Scale (MGCS) somatic score results

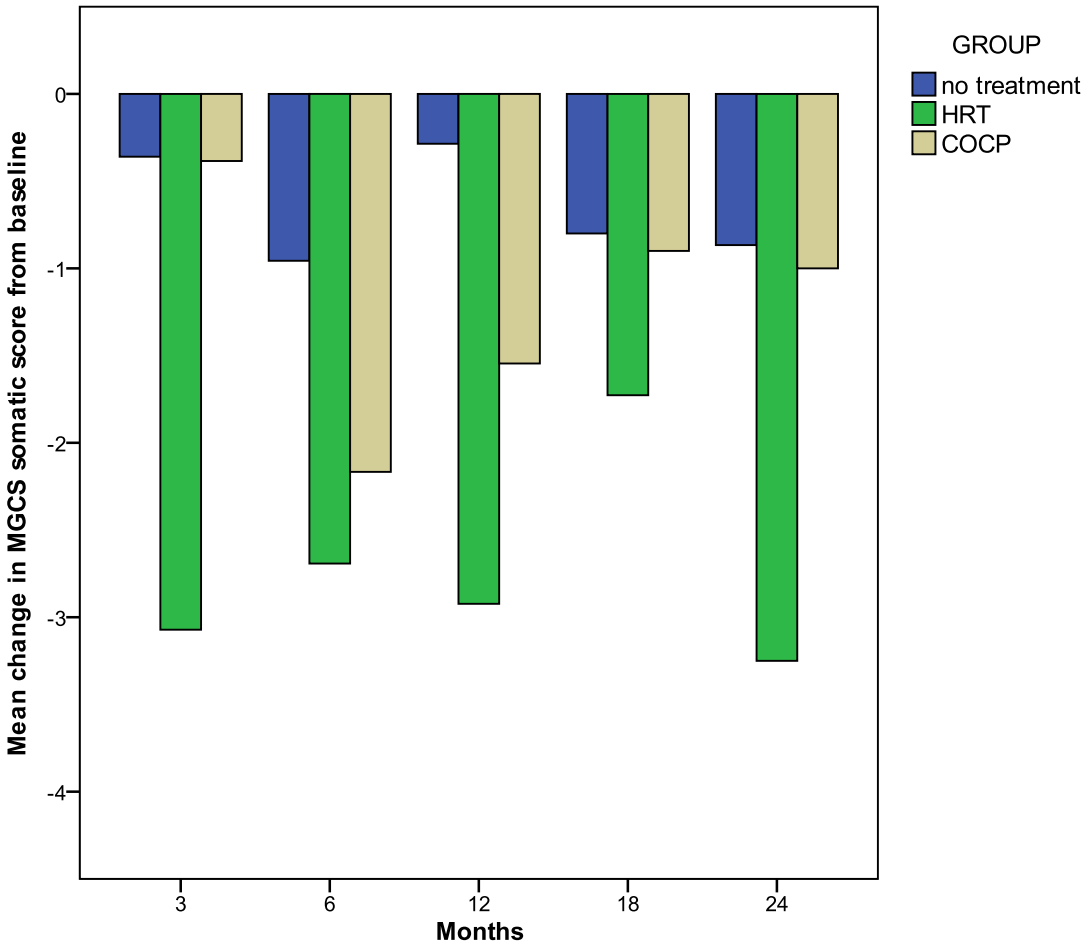


Figure 35 Changes from baseline in Modified Greene Climacteric Scale (MGCS) somatic mean score over 24 months showing all available data at each time-point

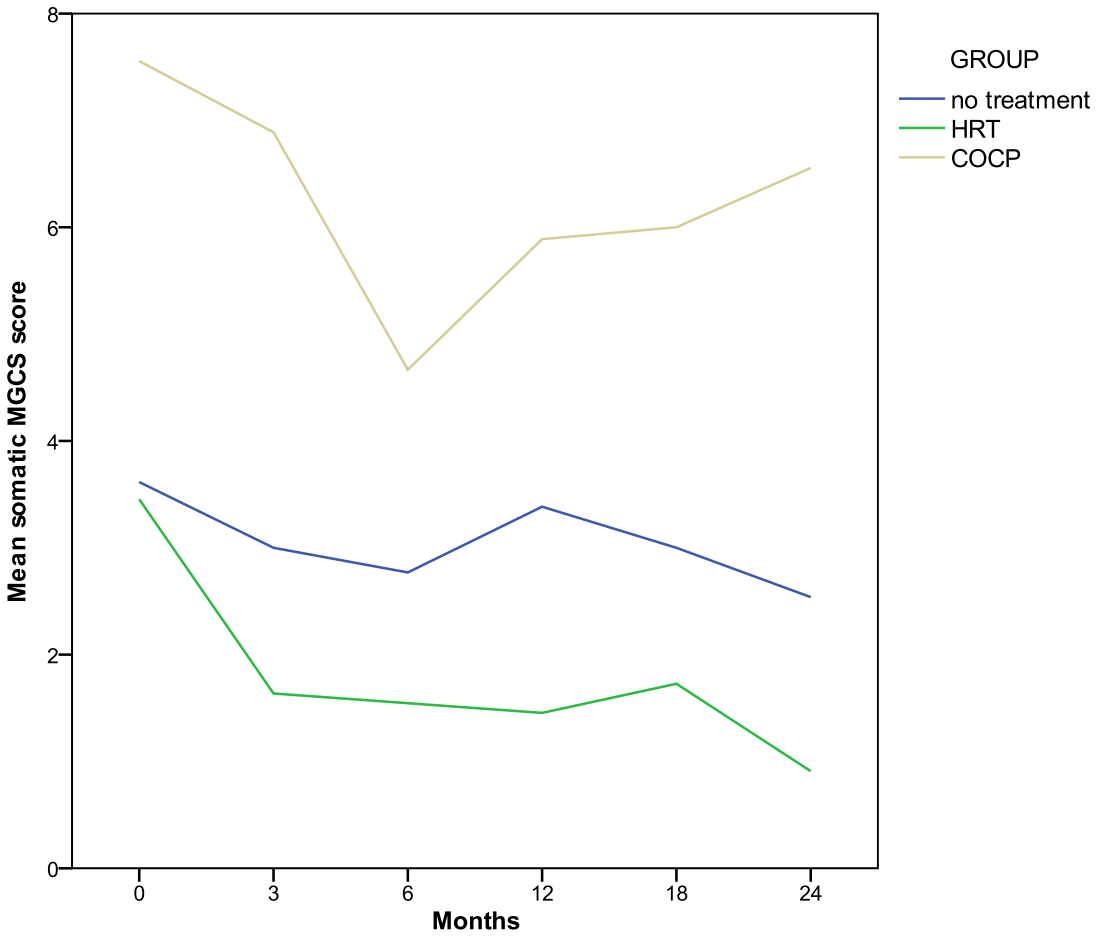


Figure 36 Changes in Modified Greene Climacteric Scale (MGCS) mean somatic score over 24 months in participants with complete data collection

4.15.5.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean MGCS somatic results	95% confidence interval of the difference	p value
3	3.0	0.9 to 5.0	0.006
6	1.2	-0.5 to 2.9	0.172
12	2.4	0.2 to 4.6	0.034
18	3.4	1.0 to 5.9	0.008
24	3.8	2.3 to 5.4	<0.001

Table 67 Comparison between HRT and COCP Modified Greene Climacteric Scale (MGCS) somatic score results. Linear regression analysis was used to adjust for baseline score.

4.15.5.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean MGCS somatic results	95% confidence interval of the difference	p value
3	-2.0	-3.3 to -0.7	0.003
6	-0.9	-2.5 to 0.7	0.249
12	-1.8	-3.4 to -0.1	0.034
18	-1.0	-2.4 to 0.4	0.152
24	-1.9	-3.6 to -0.2	0.031

Table 68 Comparison between HRT and no treatment Modified Greene Climacteric Scale (MGCS) somatic score results. Linear regression analysis was used to adjust for baseline score.

4.15.5.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean MGCS somatic results	95% confidence interval of the difference	p value
3	0.3	-0.5 to 1.2	0.425
6	0.2	-0.7 to 1.0	0.705
12	0.2	-0.9 to 1.3	0.687
18	0.9	-0.4 to 2.1	0.155
24	0.6	-0.4 to 1.6	0.246

Table 69 Comparison between COCP and no treatment Modified Greene Climacteric Scale (MGCS) somatic score results. Linear regression analysis was used to adjust for baseline score.

4.15.6 Vasomotor score

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline MGCS vasomotor score	15	2.3 (2.4)	15	3.4 (2.0)	29	1.5 (1.8)
Change from baseline to 3 months	14	-2.0 (2.4)	13	-1.8 (2.7)	25	+0.4 (1.5)
Change from baseline to 6 months	13	-1.5 (2.2)	12	-2.3 (2.8)	23	+0.5 (1.6)
Change from baseline to 12 months	13	-1.8 (2.6)	11	-2.3 (2.7)	21	+0.6 (1.4)
Change from baseline to 18 months	11	-1.5 (2.5)	10	-1.9 (3.0)	15	+0.3 (2.2)
Change from baseline to 24 months	12	-1.7 (2.5)	9	-1.8 (2.7)	15	-0.3 (2.2)

Table 70 Modified Greene Climacteric Scale (MGCS) vasomotor score results

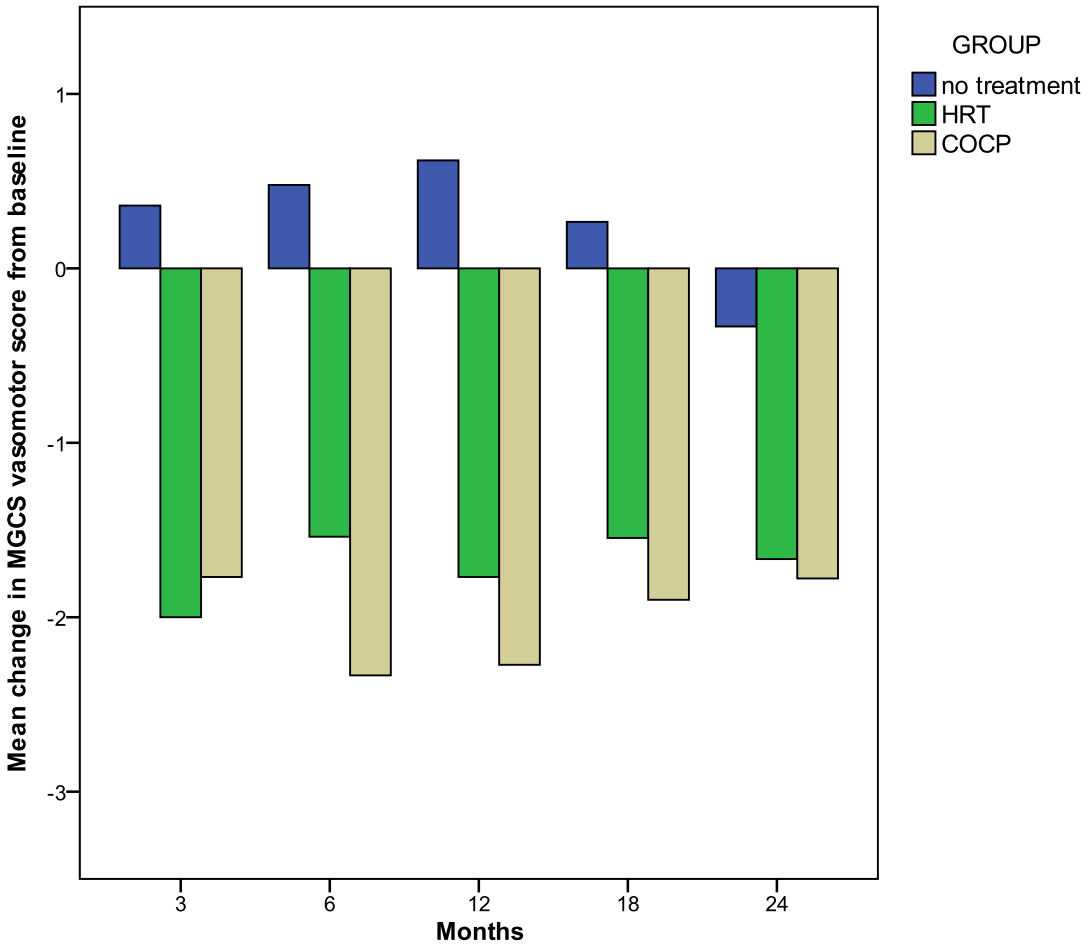


Figure 37 Changes from baseline in Modified Greene Climacteric Scale (MGCS) vasomotor mean score over 24 months showing all available data at each time-point

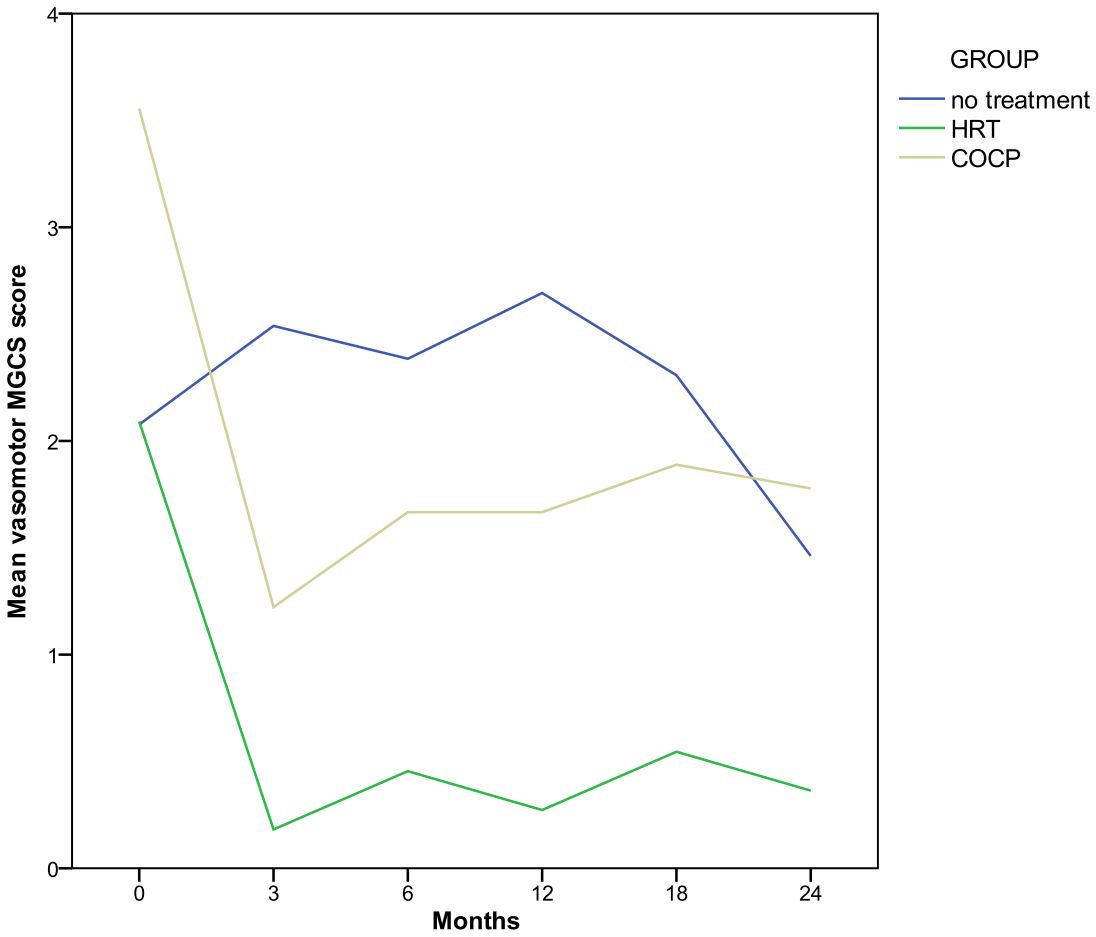


Figure 38 Changes in Modified Greene Climacteric Scale (MGCS) mean vasomotor score over 24 months in participants with complete data collection

4.15.6.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean MGCS vasomotor results	95% confidence interval of the difference	p value
3	1.1	0.0 to 2.1	0.048
6	0.4	-0.9 to 1.7	0.558
12	0.9	-0.5 to 2.3	0.209
18	1.0	-0.7 to 2.7	0.240
24	1.1	-0.2 to 2.5	0.094

Table 71 Comparison between HRT and COCP Modified Greene Climacteric Scale (MGCS) vasomotor score results. Linear regression analysis was used to adjust for baseline score.

4.15.6.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean MGCS vasomotor results	95% confidence interval of the difference	p value
3	-1.9	-2.9 to -0.9	<0.001
6	-1.6	-2.6 to -0.5	0.004
12	-1.9	-2.8 to -1.0	<0.001
18	-1.6	-2.7 to -0.6	0.005
24	-1.2	-2.5 to 0.1	0.068

Table 72 Comparison between HRT and no treatment Modified Greene Climacteric Scale (MGCS) vasomotor score results. Linear regression analysis was used to adjust for baseline score.

4.15.6.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean MGCS vasomotor results	95% confidence interval of the difference	p value
3	-0.6	-1.3 to 0.0	0.061
6	-0.8	-1.5 to -0.0	0.039
12	-0.8	-1.5 to -0.1	0.033
18	-0.5	-1.4 to 0.2	0.189
24	-0.2	-1.2 to 0.7	0.623

Table 73 Comparison between COCP and no treatment Modified Greene Climacteric Scale (MGCS) vasomotor score results. Linear regression analysis was used to adjust for baseline score.

4.15.7 Sexual Dysfunction score

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline MGCS sexual dysfunction score	15	1.4 (1.2)	15	1.7 (1.2)	29	1.1 (0.9)
Change from baseline to 3 months	14	-0.6 (1.0)	13	-0.8 (1.3)	25	0.0 (0.6)
Change from baseline to 6 months	13	-0.7 (1.1)	12	-0.8 (1.3)	23	-0.1 (0.7)
Change from baseline to 12 months	13	-0.8 (1.1)	11	-0.3 (1.6)	21	+0.1 (0.9)
Change from baseline to 18 months	11	-0.4 (1.1)	10	-0.3 (1.5)	15	0.0 (0.5)
Change from baseline to 24 months	12	-0.7 (1.2)	9	-0.8 (1.2)	15	0.0 (0.8)

Table74 Modified Greene Climacteric Scale (MGCS) sexual dysfunction score results

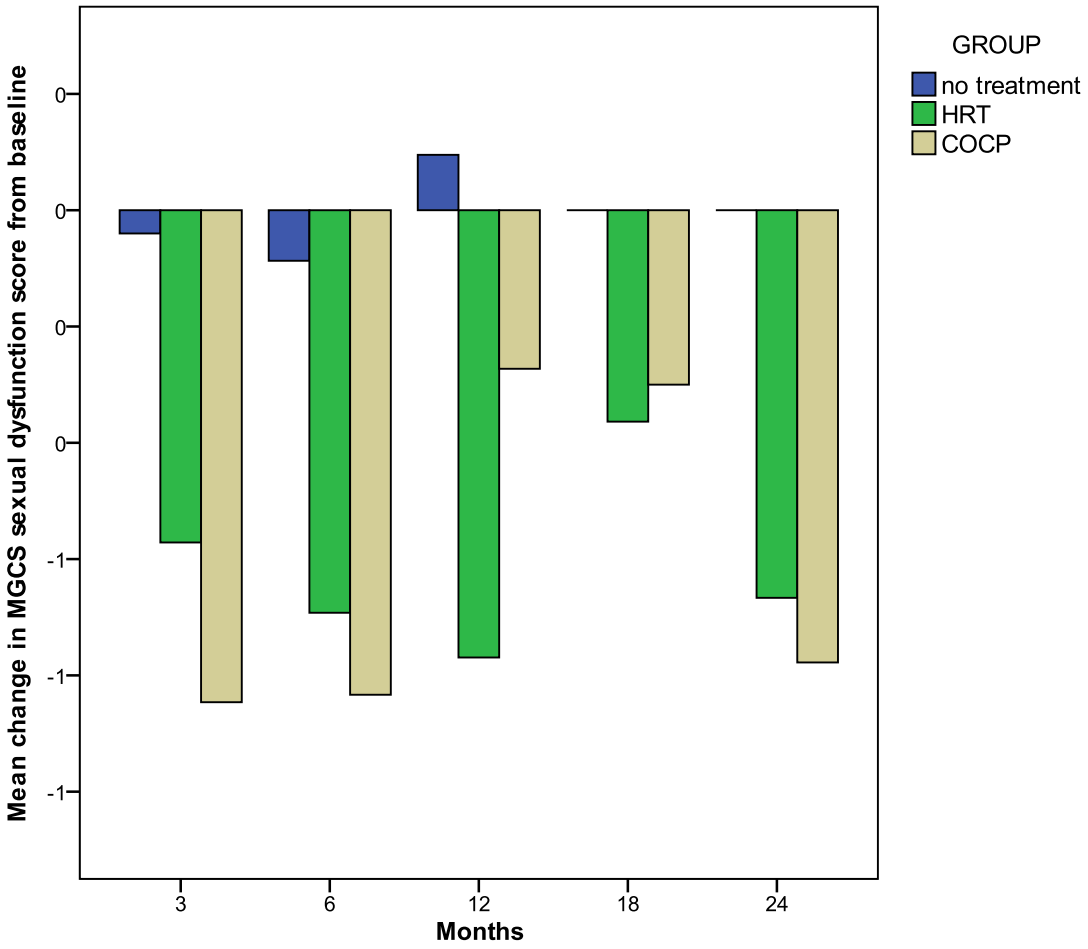


Figure 39 Changes from baseline in Modified Greene Climacteric Scale (MGCS) sexual dysfunction mean score over 24 months showing all available data at each time-point

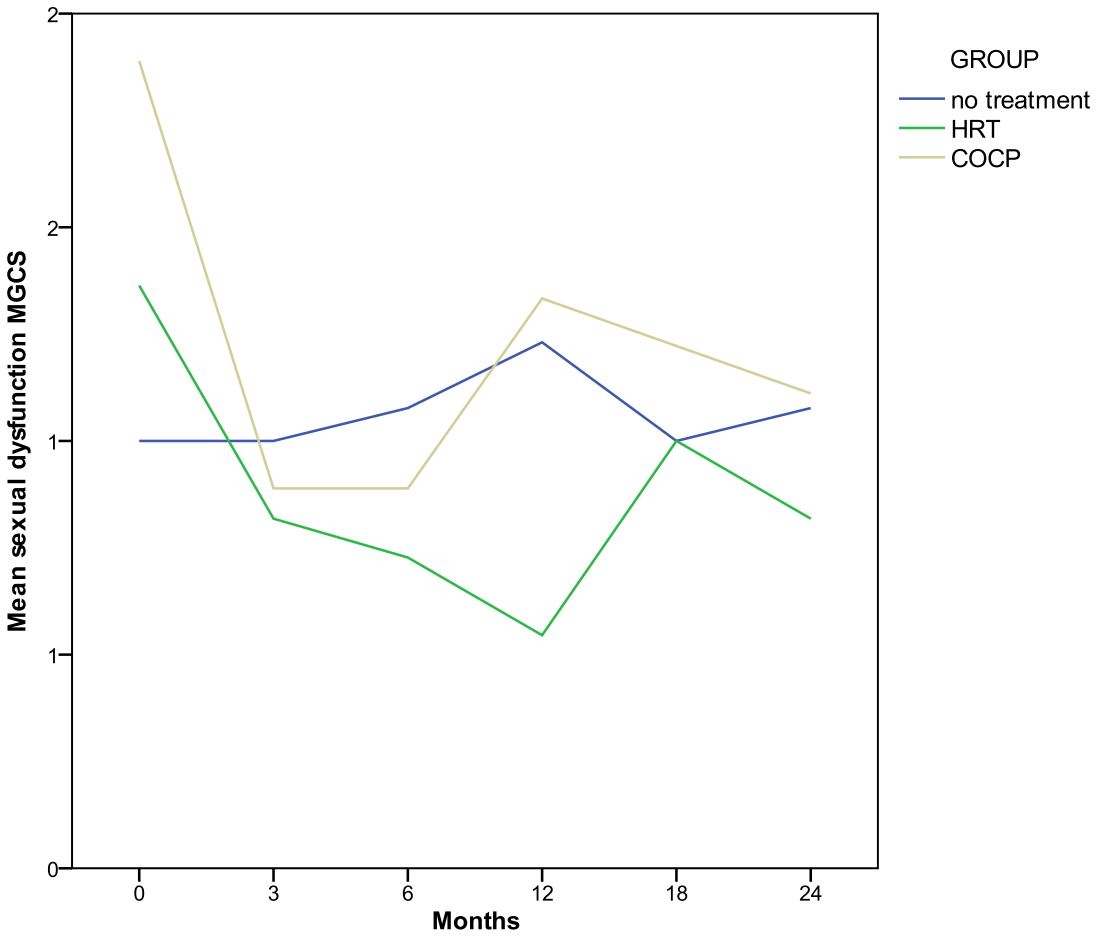


Figure 40 Changes in Modified Greene Climacteric Scale (MGCS) mean sexual dysfunction score over 24 months in participants with complete data collection

4.15.7.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean MGCS sexual dysfunction results	95% confidence interval of the difference	p value
3	-0.1	-0.7 to 0.5	0.673
6	0.0	-0.7 to 0.7	0.990
12	0.6	-0.3 to 1.4	0.213
18	0.2	-0.9 to 1.3	0.665
24	0.1	-0.8 to 1.1	0.752

Table 75 Comparison between HRT and COCP Modified Greene Climacteric Scale (MGCS) sexual dysfunction score results. Linear regression analysis was used to adjust for baseline score.

4.15.7.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean MGCS sexual dysfunction results	95% confidence interval of the difference	p value
3	-0.4	-0.9 to 0.1	0.103
6	-0.3	-0.9 to 0.2	0.220
12	-0.6	-1.2 to -0.0	0.049
18	-0.3	-0.9 to 0.4	0.436
24	-0.5	-1.1 to 0.2	0.147

Table 76 Comparison between HRT and no treatment Modified Greene Climacteric Scale (MGCS) sexual dysfunction score results. Linear regression analysis was used to adjust for baseline score.

4.15.7.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean MGCS sexual dysfunction results	95% confidence interval of the difference	p value
3	-0.3	-0.5 to -0.0	0.024
6	-0.1	-0.4 to 0.2	0.373
12	-0.0	-0.4 to 0.4	0.912
18	0.0	-0.4 to 0.4	0.974
24	-0.2	-0.7 to 0.2	0.234

Table 77 Comparison between COCP and no treatment Modified Greene Climacteric Scale (MGCS) sexual dysfunction score results. Linear regression analysis was used to adjust for baseline score.

4.16 Menopause Symptoms Treatment Satisfaction Questionnaire

The Menopause Symptoms Treatment Satisfaction Questionnaire (MS-TSQ) was completed by women in the HRT and COCP groups at 3, 6, 12, 18 and 24 months. There is no baseline score, because it assesses satisfaction with treatment. The results therefore need to be interpreted with caution because of the high drop-out rate, especially in the COCP group, and because we know from the Modified Greene Climacteric Scale results that the women who completed the trial in the COCP group had higher baseline symptom scores than those who completed the trial in the HRT group.

Comparison between the groups revealed a significant difference at 3 months (p=0.003) in favour of HRT. This was not seen at 6 months, which fits with other questionnaires used in this trial (Modified Greene Climacteric Score, Patient Health Questionnaire-9) in which the COCP group took longer to experience symptom reduction. At 12 months satisfaction with the COCP dropped and the difference between HRT and COCP is again significant in favour of HRT (p=0.040) but this reduces at 18 and 24 months, possibly because of a decreasing number of women in the trial.

	HRT		COCP	
	Number of women	Mean (SD)	Number of women	Mean (SD)
3 months	12	82 (12)	13	65 (13)
6 months	13	76 (25)	12	70 (11)
12 months	13	76 (23)	11	57 (19)
18 months	11	79 (13)	10	66 (17)
24 months	12	80 (15)	8	72 (18)

Table 78 Menopause Symptoms Treatment Satisfaction Questionnaire (MS-TSQ) results

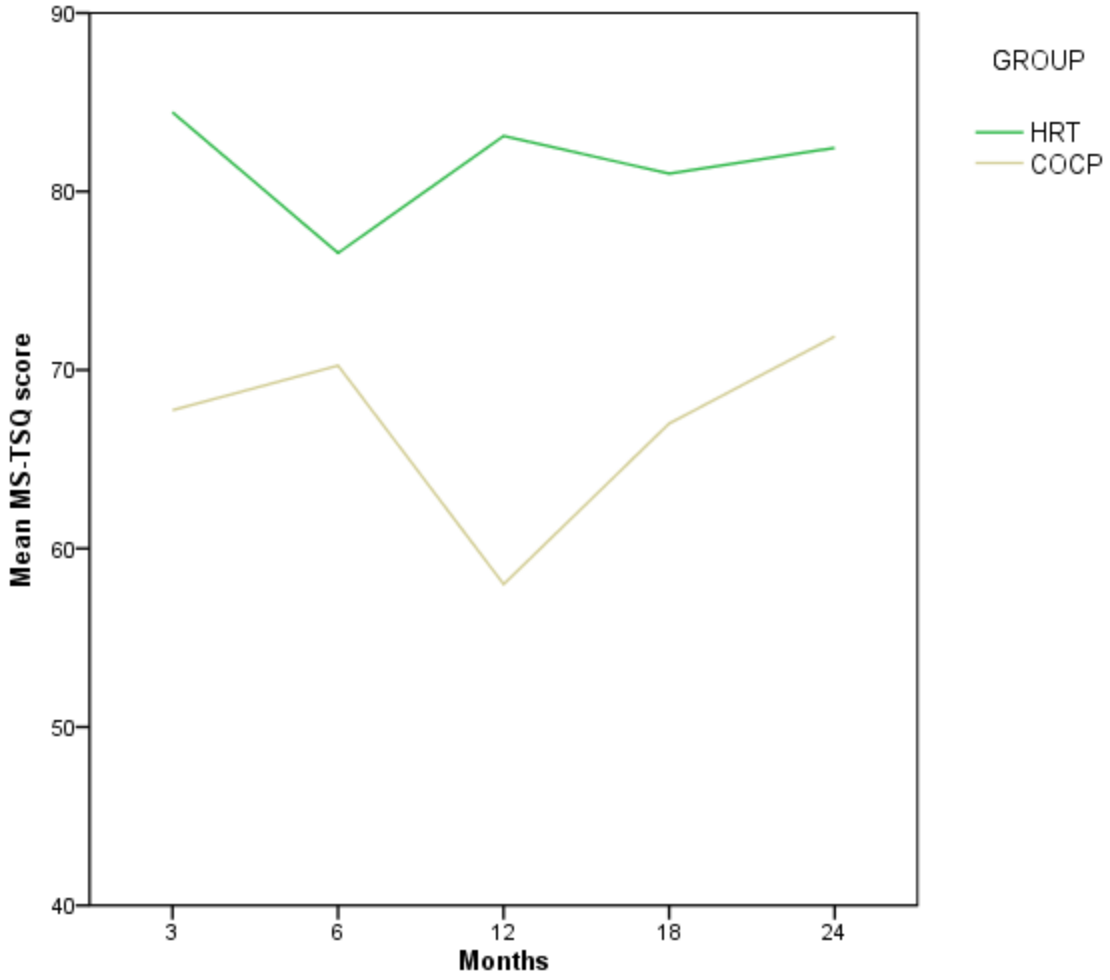


Figure 41 Mean Menopause Symptoms Treatment Satisfaction Questionnaire (MS-TSQ) score over 24 months in participants with complete data collection

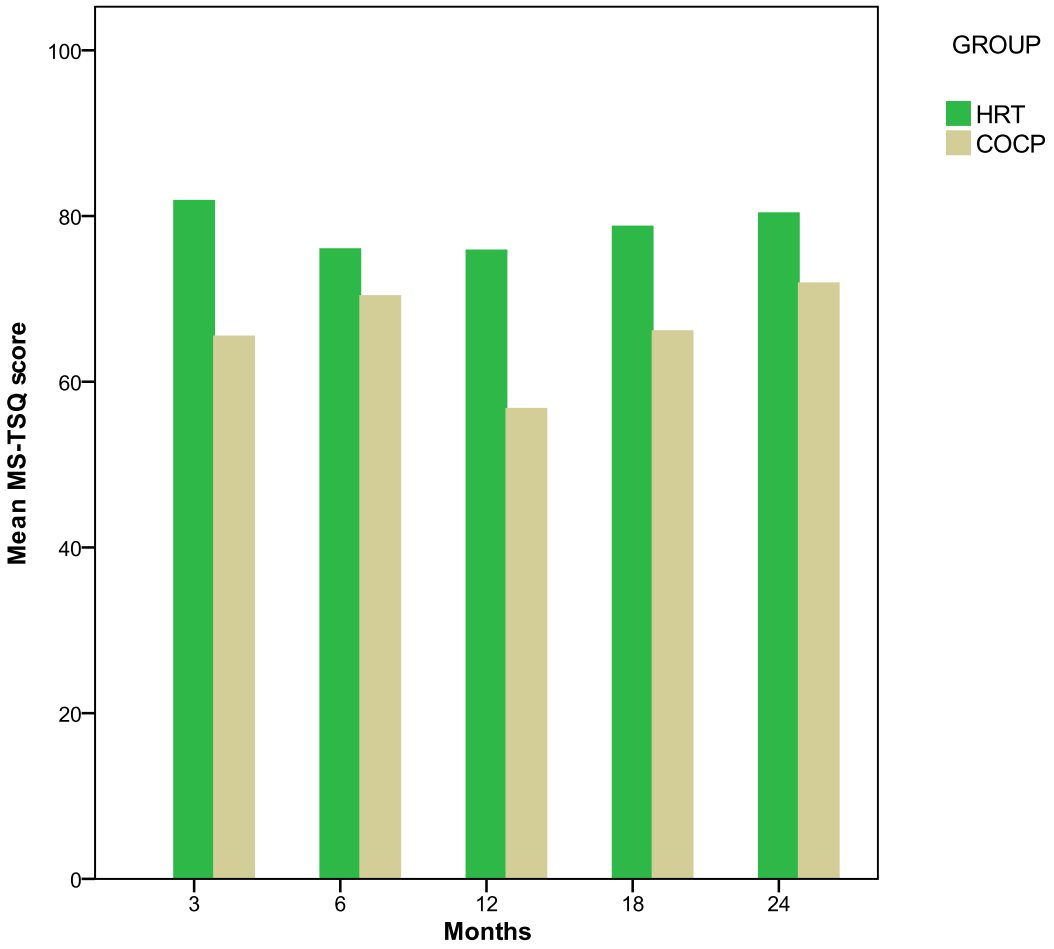


Figure 42 Mean Menopause Symptoms Treatment Satisfaction Questionnaire (MS-TSQ) score over 24 months showing all available data at each time-point

4.16.1 Comparison between groups

Months	COCP minus HRT mean MS-TSQ score	95% confidence interval of the difference	p value
3	-16.4	-26.7 to -6.1	0.003
6	-5.7	-22.0 to +10.7	0.480
12	-19.1	-37.3 to -0.98	0.040
18	-12.6	-26.2 to +0.9	0.066
24	-8.5	-23.8 to +6.8	0.260

Table 79 Comparison between HRT and COCP Menopause Symptoms Treatment Satisfaction Questionnaire (MS-TSQ) score results. Independent sample t-tests were used to compare mean scores.

4.17 Brief Profile of Female Sexual Function

There were more missing data from this questionnaire than the others due to some women not feeling able to complete the questionnaire, largely due to not being sexually active. Fig 43 shows that the Brief Profile of Female Sexual Function (BPFSF) scores in the no treatment and COCP groups remained relatively unchanged throughout the trial, whereas in the HRT group an increase was seen from 3 months and was maintained at a similar level over the duration of the trial. Fig 44, which illustrates the scores from subjects with complete data collection, shows that from almost identical baseline scores in the HRT and COCP groups, the score in the HRT group increased at 3 months and then remained relatively stable for the duration of the trial. However, when the HRT and COCP groups were compared formally, although there was a trend in favour of HRT increasing the BPFSF score to a greater extent, the differences did not reach a significance level of <0.05. Comparison between HRT and no treatment revealed significant differences at most time-points; a reduction in significance level at 24 months was likely due to decreased numbers. In contrast, there were no significant differences between the COCP and no treatment groups.

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline BPFSF score	13	18.1 (7.3)	13	18.0 (9.2)	23	20.4 (7.7)
Change from baseline to 3 months	13	+5.2 (6.2)	11	+1.2 (7.3)	19	+0.6 (4.2)
Change from baseline to 6 months	12	+5.8 (6.3)	11	+1.6 (7.0)	17	+-.1 (5.1)
Change from baseline to 12 months	12	+3.9 (4.9)	10	-1.3 (7.3)	15	-1.7 (5.6)
Change from baseline to 18 months	11	+3.5 (4.9)	9	-1.3 (9.9)	12	-0.8 (4.3)
Change from baseline to 24 months	11	+5.0 (7.0)	8	-0.1 (9.4)	13	+0.4 (6.2)

Table 80 Brief Profile of Female Sexual Function (BPFSF) results

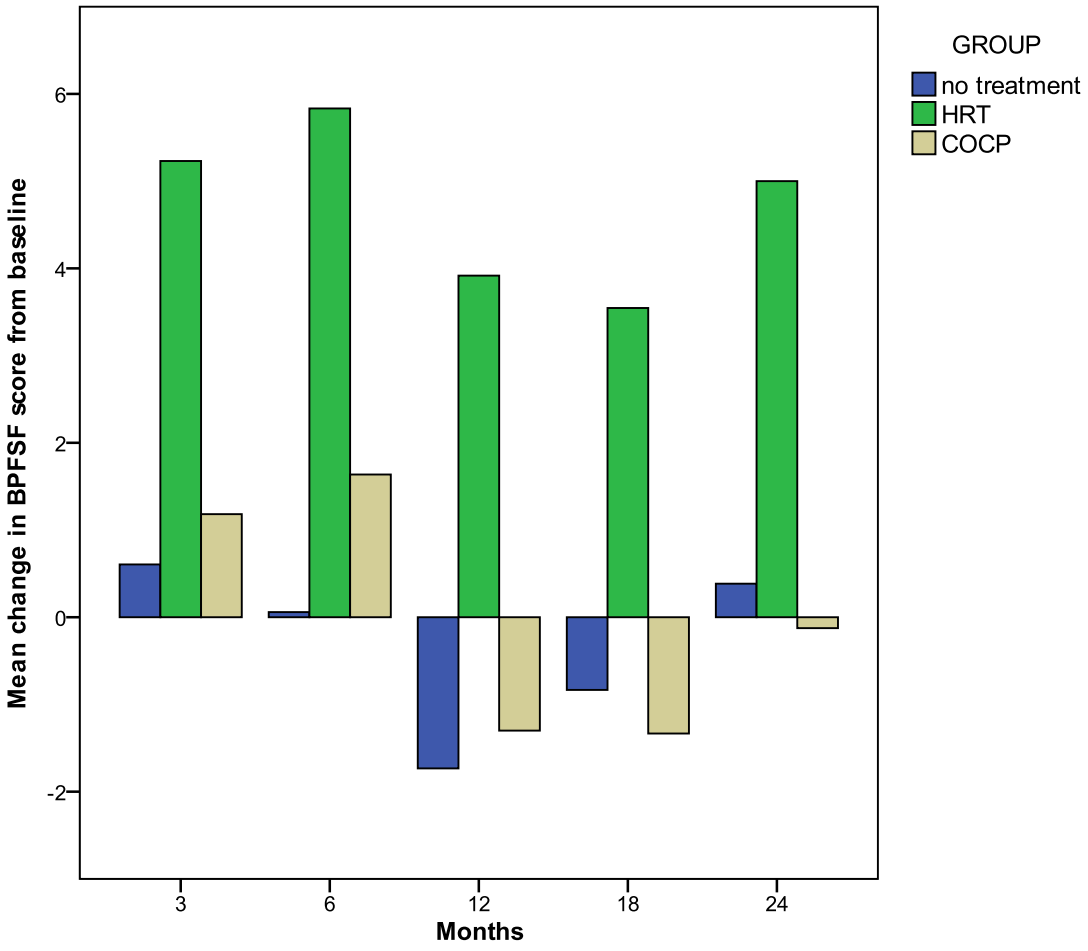


Figure 43 Changes from baseline in Brief Profile of Female Sexual Function (BPFSF) mean score over 24 months showing all available data at each time-point

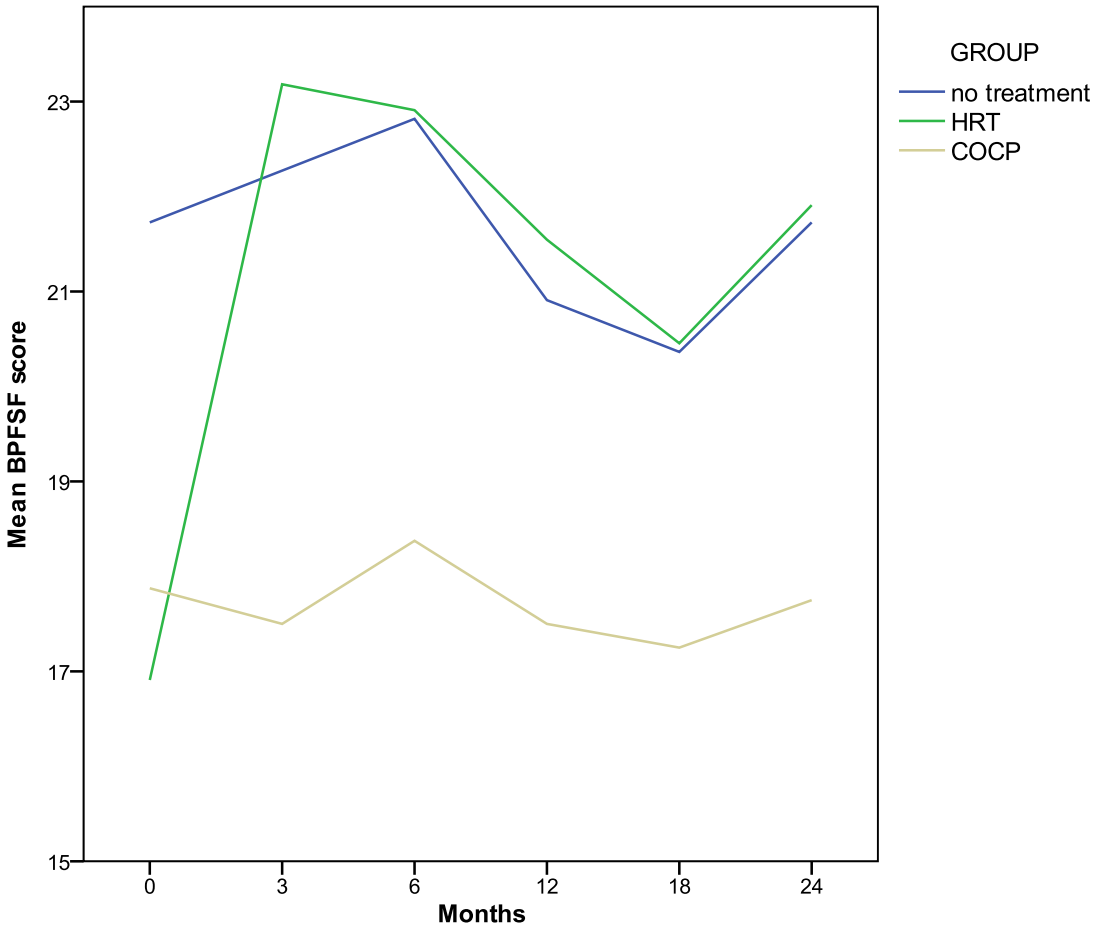


Figure 44 Changes in mean Brief Profile of Female Sexual Function (BPFSF) score over 24 months in participants with complete data collection

4.17.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean BPFSF score	95% confidence interval of the difference	p value
3	-4.1	-9.0 to 0.8	0.097
6	-3.9	-8.3 to 0.5	0.081
12	-4.4	-9.3 to 0.5	0.074
18	-4.3	-10.9 to 2.4	0.193
24	-4.7	-12.0 to 2.7	0.195

Table 81 Comparison between HRT and COCP Brief Profile of Female Sexual Function (BPFSF) score results. Linear regression analysis was used to adjust for baseline score

4.17.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean BPFSF score	95% confidence interval of the difference	p value
3	4.1	0.3 to 7.9	0.036
6	4.3	-0.2 to 8.8	0.060
12	4.4	0.4 to 8.3	0.032
18	4.5	0.1 to 8.8	0.045
24	2.8	-2.1 to 7.7	0.254

Table 82 Comparison between HRT and no treatment Brief Profile of Female Sexual Function (BPFSF) score results. Linear regression analysis was used to adjust for baseline score

4.17.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean BPFSF score	95% confidence interval of the difference	p value
3	-0.2	-2.2 to 1.8	0.860
6	-0.1	-2.2 to 2.0	0.900
12	-0.2	-2.5 to 2.1	0.886
18	-1.1	-4.1 to 1.9	0.447
24	-1.0	-4.0 to 1.9	0.459

Table 83 Comparison between COCP and no treatment Brief Profile of Female Sexual Function (BPFSF) score results. Linear regression analysis was used to adjust for baseline score

4.18 Short Form-36

Results are shown as mean (standard deviation) when normally distributed and median (25, 75%) when not normally distributed. Comparisons between groups were carried out using linear regression with adjustment for baseline score in the case of normally distributed data, or the Mann-Whitney U test where data are not normally distributed. In several of the domains, there are no changes from baseline over the duration of the trial (especially in the HRT and no treatment groups) and graphs are not shown in these cases. In both the HRT and the COCP groups, there were a few women whose scores rose dramatically in most domains from 3 months, indicating that treatment can have huge positive effects on quality of life in some women. However, due to the distribution of the data and the fact that this did not occur in most women on treatment, this effect is not appreciated from the data displayed below.

4.18.1 Physical functioning

	HRT		COCP		No treatment	
	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)
Baseline SF-36 physical functioning score	15	100.0 (85.0, 100.0)	15	90.0 (65.0, 100.0)	29	100.0 (85.0, 100.0)
Change from baseline to 3 months	14	0.0 (0.0, 6.3)	13	0.0 (-5.0, 21.0)	25	0.0 (-2.5, 0.0)
Change from baseline to 6 months	13	0.0 (0.0, 7.5)	12	+5.0 (1.3, 20.0)	23	0.0 (-5.0, 5.0)
Change from baseline to 12 months	13	0.0 (0.0, 7.5)	11	+5.0 (0.0, 20.0)	20	0.0 (-3.8, 0.0)
Change from baseline to 18 months	11	0.0 (0.0, 5.0)	10	0.0 (-1.3, 25.0)	15	0.0 (-5.0, 5.0)
Change from baseline to 24 months	12	0.0 (0.0, 12.5)	9	+5.0 (0.0, 27.5)	15	0.0 (0.0. 0.0)

Table 84 Short Form-36 (SF-36) physical functioning results

Mann-Whitney U tests revealed a significant difference between the changes in score in the HRT and COCP groups at 6 months but not at any other time-points (p values 0.513, 0.048, 0.306, 0.526 and 0.429 at 3, 6, 12, 18 and 24 months). This raises the suspicion that the p value of <0.05 at 6 months was due to chance. There were no significant differences between the changes in the HRT and no treatment groups (p values 0.054, 0.267, 0.190, 0.784 and 0.261 at 3, 6, 12, 18 and 24 months). Comparison between the COCP and no treatment groups revealed significant differences at 6 and 12 months (p values 0.583, 0.005, 0.035, 0.388 and 0.097 at 3, 6, 12, 18 and 24 months).

4.18.2 Role-Physical

	HRT		COCP		No treatment	
	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)
Baseline SF-36 role-physical score	15	94.0 (63.0, 100.0)	15	63.0 (50.0, 100.0)	29	100.0 (84.5, 100.0)
Change from baseline to 3 months	14	+3.0 (0.0, 12.0)	13	+12.0 (0.0, 35.0)	25	0.0 (-12.0, 0.0)
Change from baseline to 6 months	13	0.0 (0.0, 18.5)	12	+22.0 (0.0, 36.3)	23	0.0 (-6.0, 6.0)
Change from baseline to 12 months	13	0.0 (0.0, 18.5)	11	0.0 (-6.0, 32.0)	20	0.0 (-4.5, 0.0)
Change from baseline to 18 months	11	0.0 (0.0, 12.0)	10	+3.0 (-7.5, 35.8)	15	0.0 (-19.0, 0.0)
Change from baseline to 24 months	12	0.0 (0.0, 10.5)	9	+12.0 (0.0, 50.5)	15	0.0 (0.0, 0.0)

Table 85 Short Form-36 (SF-36) role-physical results

Mann-Whitney U tests did not show any significant differences between the changes in score in the HRT and COCP groups at any time-point (p values 0.240, 0.143, 0.904, 0.941 and 0.128 at 3, 6, 12, 18 and 24 months). There were significant differences between the changes in the HRT and no treatment groups at 3 and 12 months but not at the other time-points (p values 0.026, 0.171, 0.017, 0.088 and 0.115 at 3, 6, 12, 18 and 24 months). Comparison between the COCP and no treatment groups revealed significant differences at 3, 6 and 24 months (p values 0.008, 0.011, 0.208, 0.263 and 0.003 at 3, 6, 12, 18 and 24 months).

4.18.3 Bodily Pain

	HRT		COCP		No treatment	
	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)
Baseline SF-36 bodily pain score	15	84.0 (61.0, 100.0)	15	52.0 (42.0, 84.0)	29	84.0 (62.0, 100.0)
Change from baseline to 3 months	14	0.0 (-16.0, 11.5)	13	0.0 (-13.5, 10.5)	25	0.0 (-10.5, 0.0)
Change from baseline to 6 months	13	0.0 (0.0, 14.0)	12	+11.0 (-7.5, 29.8)	23	0.0 (-16.0, 12.0)
Change from baseline to 12 months	13	0.0 (0.0, 17.0)	11	0.0 (-10.0, 16.0)	20	0.0 (-15.0, 0.0)
Change from baseline to 18 months	11	0.0 (0.0, 16.0)	10	+10.0 (-2.5, 20.0)	15	0.0 (-11.0, 21)
Change from baseline to 24 months	12	0.0 (0.0, 22.0)	9	+2.0 (-5.0, 20.5)	15	0.0 (0.0, 22.0)

Table 86 Short Form-36 (SF-36) bodily pain results

Mann-Whitney U tests did not show any significant differences between the changes in score in any of the groups at any time-point (p values for HRT versus COCP group 0.678, 0.387, 0.552, 0.826 and 0.913 at 3, 6, 12, 18 and 24 months; p values for HRT vs no treatment group 0.834, 0.398, 0.169, 0.323 and 0.772 at 3, 6, 12, 18 and 24 months; p values for COCP vs no treatment group 0.473, 0.082, 0.566, 0.233 and 0.665 at 3, 6, 12, 18 and 24 months).

4.18.4 General Health

Linear regression with adjustment for baseline score did not show any differences between the groups (tables 88 to 90).

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline SF-36 general health score	15	65.6 (27.1)	15	61.3 (24.5)	29	63.9 (24.6)
Change from baseline to 3 months	14	+4.4 (8.5)	13	+6.5 (14.6)	25	+1.6 (15.8)
Change from baseline to 6 months	13	+5.2 (8.7)	12	+7.0 (14.0)	23	+4.2 (9.2)
Change from baseline to 12 months	13	+5.9 (9.8)	11	+5.0 (15.0)	20	+1.1 (16.1)
Change from baseline to 18 months	11	+3.8 (7.7)	10	+7.5 (13.0)	15	+1.1 (14.4)
Change from baseline to 24 months	12	+4.5 (8.6)	9	+5.0 (16.0)	15	+6.2 (14.1)

Table 87 Short Form-36 (SF-36) general health results

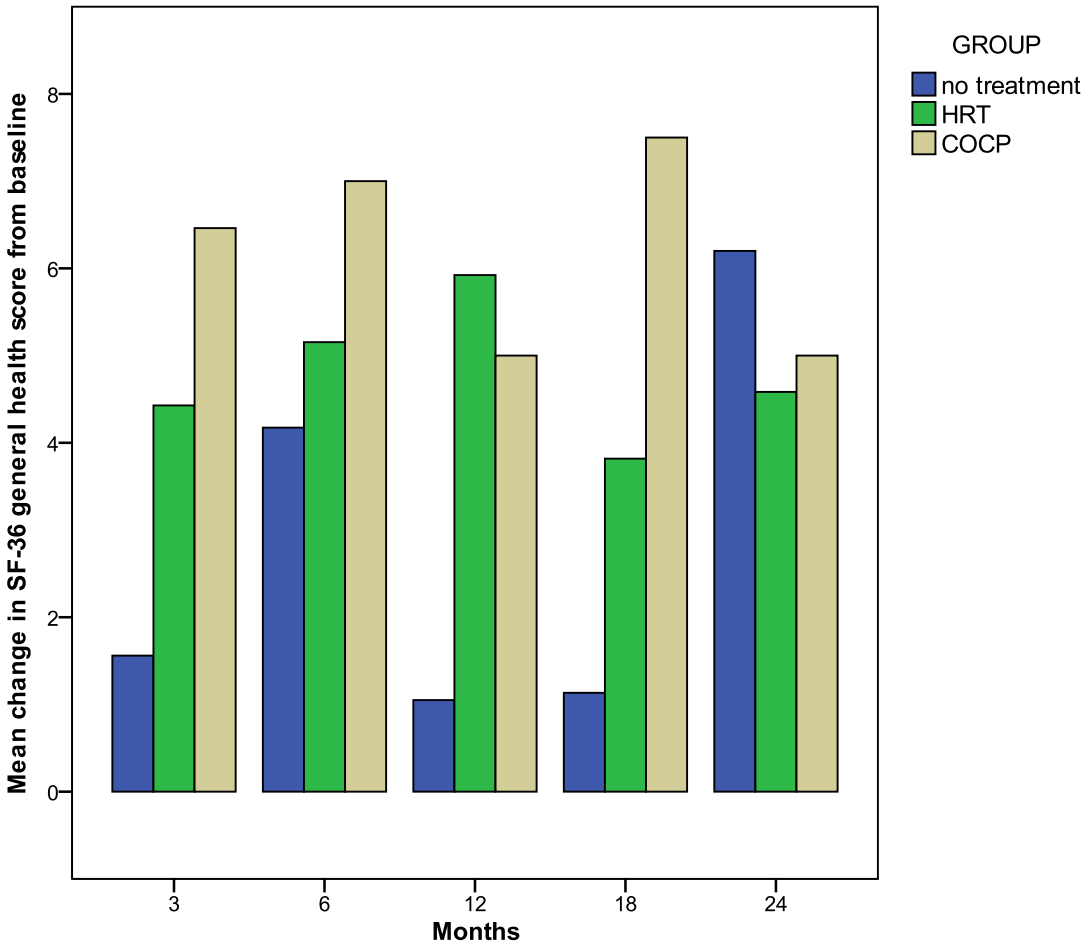


Figure 45 Changes from baseline in Short Form-36 (SF-36) General Health mean score over 24 months showing all available data at each time-point

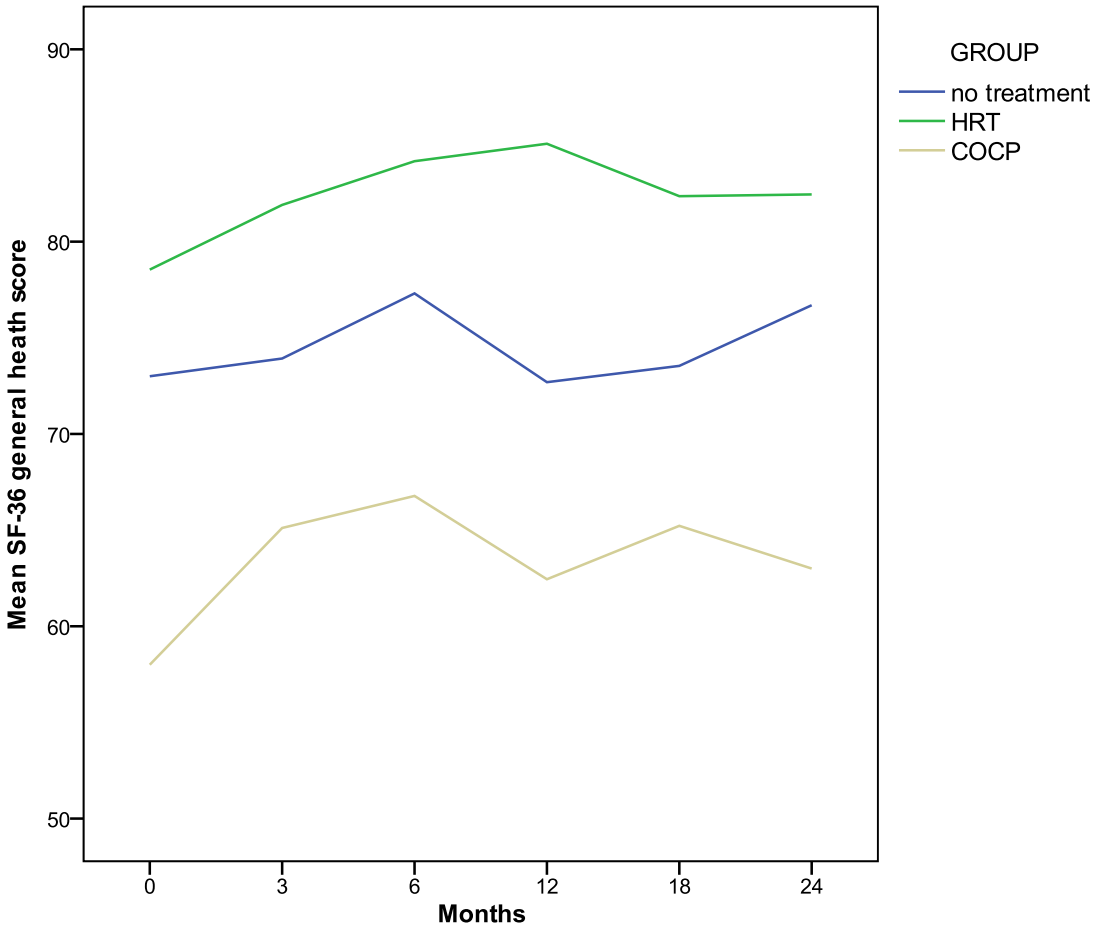


Figure 46 Changes in mean Short Form-36 (SF-36) general health score over 24 months in participants with complete data collection

4.18.4.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean SF-36 general health score	95% confidence interval of the difference	p value
3	0.19	-7.8 to 8.2	0.962
6	0.94	-8.7 to 10.5	0.841
12	-2.8	-13.2 to 7.7	0.588
18	-2.0	-10.5 to 6.4	0.621
24	-4.5	-14.6 to 5.6	0.359

Table 88 Comparison between HRT and COCP Short Form-36 (SF-36) general health score results. Linear regression analysis was used to adjust for baseline score

4.18.4.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean SF-36 general health score	95% confidence interval of the difference	p value
3	3.3	-5.9 to 12.6	0.469
6	0.7	-5.7 to 7.1	0.822
12	5.0	-5.2 to 15.2	0.325
18	3.0	-7.5 to 13.5	0.559
24	-0.8	-10.4 to 8.8	0.862

Table 89 Comparison between HRT and no treatment Short Form-36 (SF-36) general health score results. Linear regression analysis was used to adjust for baseline score

4.18.4.3 Comparison between COCPand no treatment groups

Months	COCP minus no treatment mean SF-36 general health score	95% confidence interval of the difference	p value
3	2.1	-3.0 to 7.3	0.402
6	1.3	-2.7 to 5.3	0.513
12	1.5	-4.7 to 7.7	0.625
18	2.5	-3.3 to 8.4	0.383
24	-1.7	-8.1 to 4.8	0.594

Table 90 Comparison between COCP and no treatment Short Form-36 (SF-36) general health score results. Linear regression analysis was used to adjust for baseline score

4.18.5 Vitality

Figs 47 and 48 show that the vitality score remained relatively constant in the no treatment group but increased in both the HRT and COCP groups. This effect appears more marked in the HRT group but this only reached a significance level of <0.05 at 3 and 18 months. Amongst women who completed the trial, those in the COCP group had a lower baseline vitality score than those in the HRT group.

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline SF-36 vitality score	15	47.1 (27.1)	15	41.4 (24.7)	29	54.2 (21.5)
Change from baseline to 3 months	14	+17.9 (15.6)	13	+4.2 (9.9)	25	+5.9 (15.2)
Change from baseline to 6 months	13	+14.4 (16.3)	12	+9.8 (10.1)	23	+4.2 (14.9)
Change from baseline to 12 months	13	+17.8 (14.7)	11	+7.6 (13.2)	20	-5.6 (21.2)
Change from baseline to 18 months	11	+13.7 (18.7)	10	+9.3 (20.7)	15	-0.9 (15.4)
Change from baseline to 24 months	12	+15.6 (18.1)	9	+8.2 (16.8)	15	+3.3 (17.2)

Table 91 Short Form-36 (SF-36) vitality results

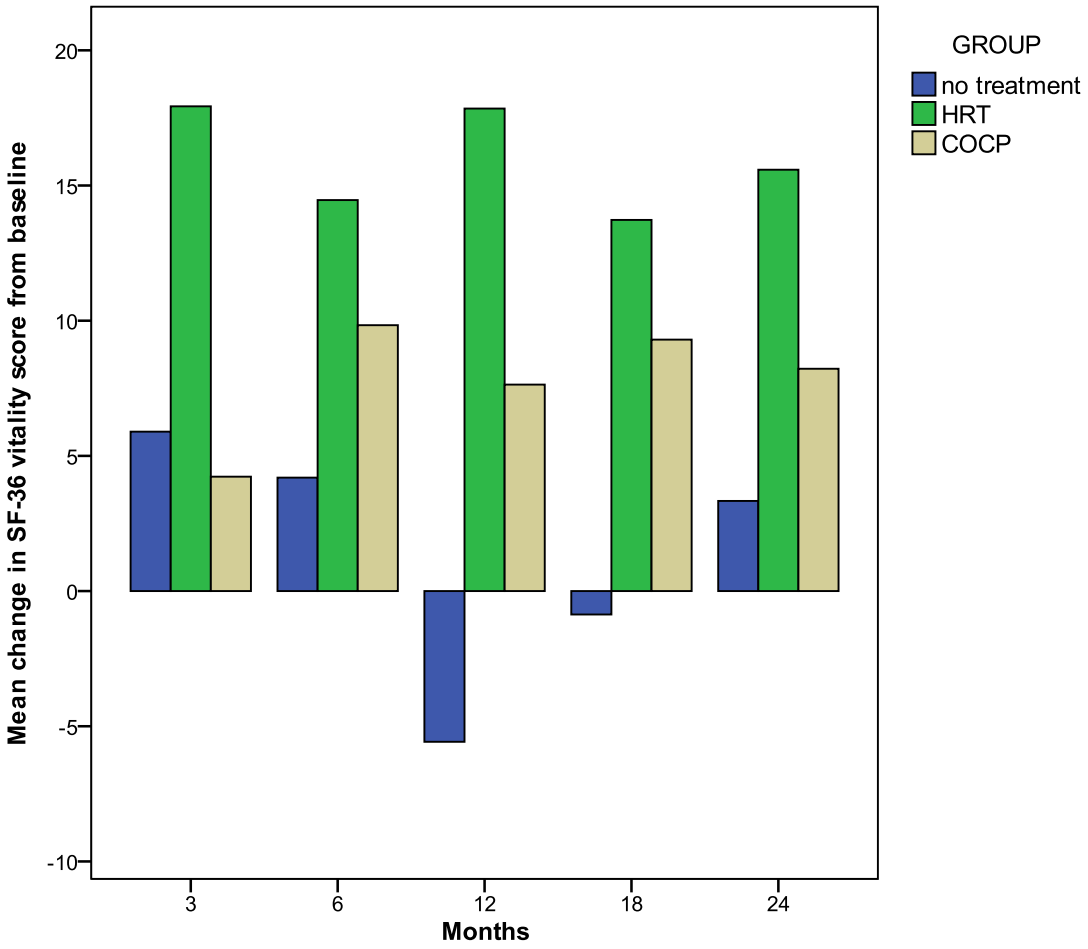


Figure 47 Changes from baseline in Short Form-36 (SF-36) vitality score over 24 months showing all available data at each time-point

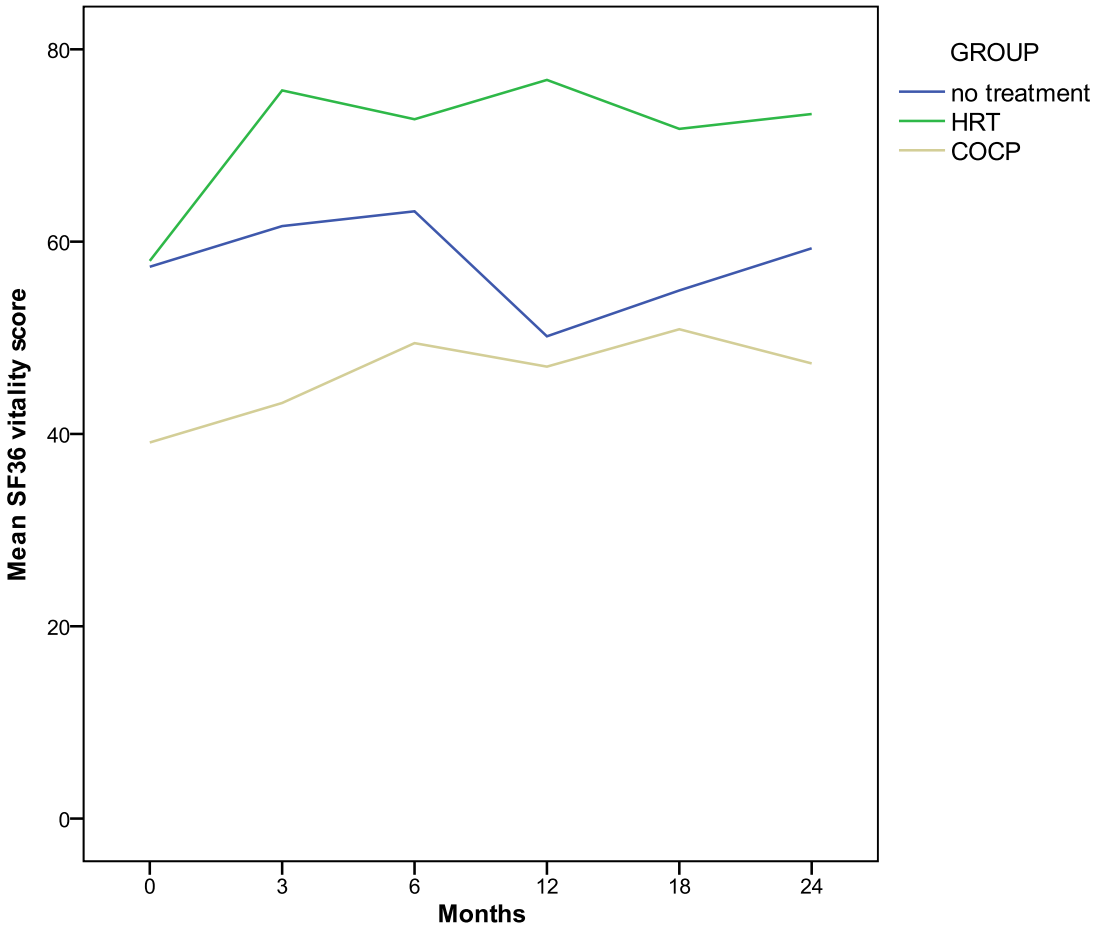


Figure 48 Changes in mean Short Form-36 (SF-36) vitality score over 24 months in participants with complete data collection

4.18.5.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean SF-36 vitality score	95% confidence interval of the difference	p value
3	-14.1	-24.4 to -3.8	0.009
6	-6.4	-16.6 to 3.8	0.207
12	-11.3	-23.6 to 1.0	0.070
18	-15.4	-28.7 to -2.1	0.025
24	-10.4	-26.9 to 6.2	0.205

Table 92 Comparison between HRT and COCP Short Form-36 (SF-36) vitality score results. Linear regression analysis was used to adjust for baseline score

4.18.5.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean SF-36 vitality score	95% confidence interval of the difference	p value
3	11.0	1.8 to 20.1	0.021
6	8.9	-1.5 to 19.2	0.091
12	22.1	8.3 to 35.9	0.003
18	15.3	2.2 to 28.4	0.024
24	12.5	-1.1 to 26.0	0.070

Table 93 Comparison between HRT and no treatment Short Form-36 (SF-36) vitality score results. Linear regression analysis was used to adjust for baseline score

4.18.5.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean SF-36 vitality score	95% confidence interval of the difference	p value
3	-1.9	-5.9 to 2.1	0.343
6	1.2	-3.5 to 6.0	0.595
12	5.2	-2.6 to 13.0	0.185
18	2.6	-5.0 to 10.1	0.490
24	1.4	-6.4 to 9.2	0.712

Table 94 Comparison between COCP and no treatment Short Form-36 (SF-36) vitality score results. Linear regression analysis was used to adjust for baseline score

4.18.6 Social Functioning

	HRT		COCP		No treatment	
	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)
Baseline SF-36 social functioning score	15	100.0 (38.0, 100.0)	15	75.0 (38.0, 75.0)	29	100.0 (69.0, 100.0)
Change from baseline to 3 months	14	0.0 (0.0, 25.0)	13	0.0 (-12.5, 25.0)	25	0.0 (-12.5, 12.5)
Change from baseline to 6 months	13	0.0 (0.0, 19.0)	12	25.0 (0.0, 34.0)	23	0.0 (-25.0, 0.0)
Change from baseline to 12 months	13	0.0 (0.0, 25.0)	11	0.0 (-13.0, 25.0)	20	0.0 (-25.0, 0.0)
Change from baseline to 18 months	11	0.0 (0.0, 25.0)	10	0.0 (-15.3, 25.0)	15	0.0 (-25.0, 0.0)
Change from baseline to 24 months	12	0.0 (0.0, 25.0)	9	0.0 (-12.5, 25.0)	15	0.0 (-12.0, 12.0)

Table 95 Short Form-36 (SF-36) social functioning results

Mann-Whitney U tests showed few significant differences between the groups (p values for HRT versus COCP group 0.495, 0.116, 0.731, 0.702 and 0.786 at 3, 6, 12, 18 and 24 months; p values for HRT vs no treatment group 0.138, 0.063, 0.039, 0.024 and 0.147 at 3, 6, 12, 18 and 24 months; p values for COCP vs no treatment group 0.642, 0.003, 0.204, 0.200 and 0.389 at 3, 6, 12, 18 and 24 months). The significant differences that were found were inconsistent and may to be due to chance.

4.18.7 Role Emotional

	HRT		COCP		No treatment	
	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)
Baseline SF-36 role-emotional score	15	92.0 (75.0, 100.0)	15	75.0 (50.0, 100.0)	29	100.0 (62.5, 100.0)
Change from baseline to 3 months	14	+4.0 (0.0, 16.3)	13	+8.0 (-8.0, 20.5)	25	0.0 (-21.0, 8.5)
Change from baseline to 6 months	13	0.0 (0.0, 13.0)	12	+21.0 (0.0, 39.0)	23	0.0 (0.0, 8.0)
Change from baseline to 12 months	13	0.0 (0.0, 21.0)	11	+17.0 (0.0, 34.0)	20	0.0 (-21.0, 0.0)
Change from baseline to 18 months	11	0.0 (0.0, 9.0)	10	+12.5 (0.0, 21.0)	15	0.0 (-9.0, 0.0)
Change from baseline to 24 months	12	0.0 (0.0, 14.3)	9	+17.0 (8.0, 46.0)	15	0.0 (0.0, 8.0)

Table 96 Short Form-36 (SF-36) role-emotional results

Mann-Whitney U tests comparing the HRT and COCP groups showed a significant difference at 24 months but not other time-points (p values 0.882, 0.108, 0.546, 0.197 and 0.035 at 3, 6, 12, 18 and 24 months. The comparison between the HRT and no treatment groups showed significant differences at 3 and 12 months only (p values 0.032, 0.217, 0.021, 0.175 and 0.249 at 3, 6, 12, 18 and 24 months). The COCP and no treatment groups were found to be significantly different at all time-points except 3 months (p values 0.149, 0.015, 0.012, 0.041 and 0.014 at 3, 6, 12, 18 and 24 months).

4.18.8 Mental Health

The mental health score improved in both the HRT and COCP groups, with larger changes seen in the HRT group. The differences between the HRT and COCP groups were not statistically significant. The HRT group had a statistically significantly increased score compared with the no treatment group at all time-points. There were no significant differences between the COCP and no treatment groups. These trends concord with the trends observed in the Patient Health Questionnaire-9 results.

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline SF-36 mental health score	15	64.7 (23.5)	15	57.7 (26.5)	29	71.2 (19.0)
Change from baseline to 3 months	14	+10.6 (10.8)	13	+4.1 (13.7)	25	-1.2 (13.9)
Change from baseline to 6 months	13	+13.4 (12.3)	12	+10.0 (19.3)	23	-2.2 (16.6)
Change from baseline to 12 months	13	+12.6 (13.5)	11	+9.2 (21.3)	20	-4.8 (18.2)
Change from baseline to 18 months	11	+8.5 (15.8)	10	+16.5 (31.2)	15	-6.7 (16.2)
Change from baseline to 24 months	12	+10.3 (12.3)	9	+16.1 (24.1)	15	+0.33 (12.0)

Table 97 Short Form-36 (SF-36) mental health results

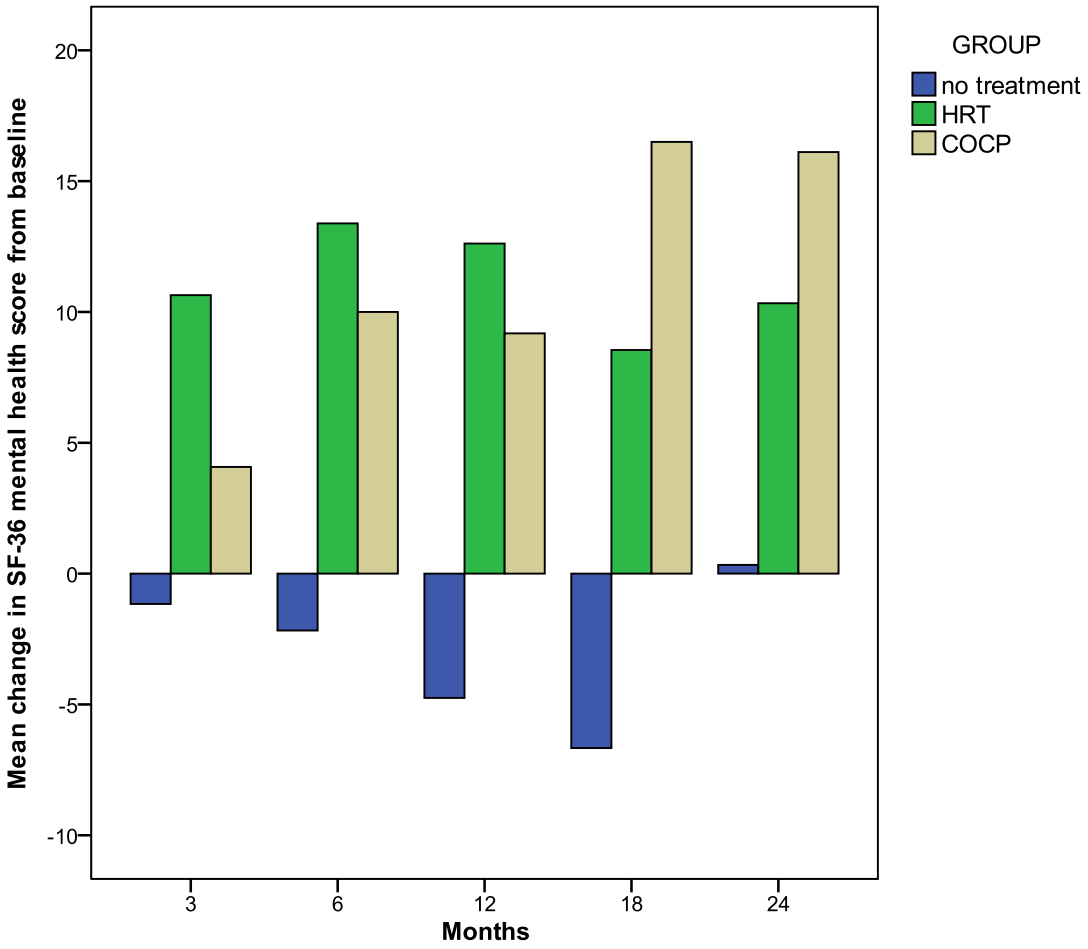


Figure 49 Changes from baseline in Short Form-36 (SF-36) mental health score over 24 months showing all available data at each time-point

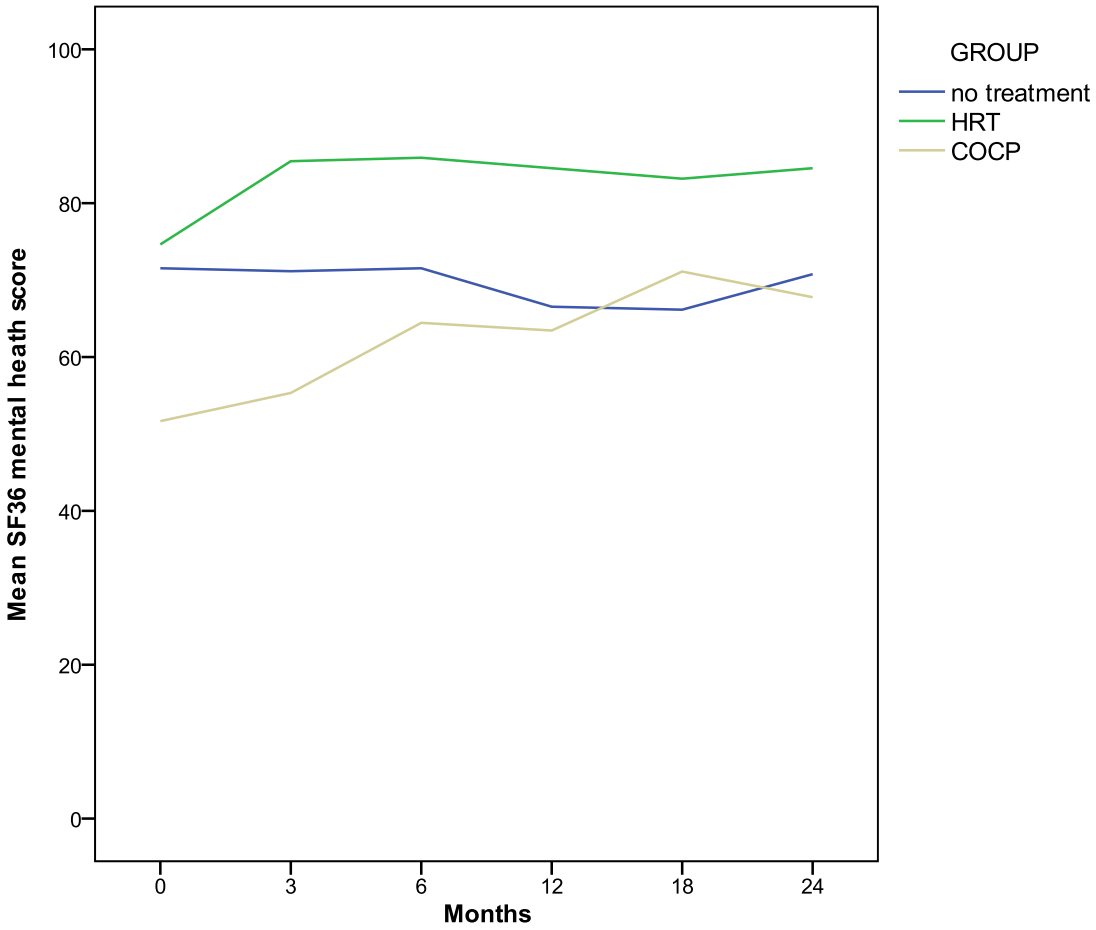


Figure 50 Changes in mean Short Form-36 (SF-36) mental health score over 24 months in participants with complete data collection

4.18.8.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean SF-36 mental health score	95% confidence interval of the difference	p value
3	-8.7	-17.5 to 0.1	0.053
6	-8.1	-18.1 to 1.8	0.103
12	-8.5	-21.4 to 4.5	0.188
18	-9.0	-24.0 to 6.0	0.225
24	-3.9	-17.3 to 9.5	0.550

Table 98 Comparison between HRT and COCP Short Form-36 (SF-36) mental health score results. Linear regression analysis was used to adjust for baseline score

4.18.8.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean SF-36 mental health score	95% confidence interval of the difference	p value
3	10.8	2.9 to 18.7	0.009
6	13.6	3.6 to 23.5	0.009
12	16.0	4.9 to 27.1	0.006
18	17.0	6.0 to 28.0	0.004
24	10.3	1.6 to 19.0	0.022

Table 99 Comparison between HRT and no treatment Short Form-36 (SF-36) mental health score results. Linear regression analysis was used to adjust for baseline score

4.18.8.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean SF-36 mental health score	95% confidence interval of the difference	p value
3	0.7	-3.7 to 5.2	0.737
6	2.8	-3.3 to 8.8	0.353
12	4.5	-3.1 to 12.0	0.236
18	5.9	-1.9 to 13.7	0.133
24	4.1	-2.6 to 10.8	0.214

Table 100 Comparison between COCP and no treatment Short Form-36 (SF-36) mental health score results. Linear regression analysis was used to adjust for baseline score

4.18.9 Physical Component Score

Fig 51 suggests higher increases in score in the COCP group compared with the other groups. However, this is not seen among the women with complete data collection (fig 52) and on formal comparison there are no differences between any of the groups.

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline SF-36 physical component score	15	51.8 (9.3)	15	49.7 (7.5)	29	52.2 (8.3)
Change from baseline to 3 months	14	+0.1 (4.5)	13	+2.9 (5.3)	25	+0.3 (5.5)
Change from baseline to 6 months	13	+1.5 (4.0)	12	+4.4 (5.4)	23	+0.3 (8.0)
Change from baseline to 12 months	13	+1.3 (4.0)	11	+2.1 (5.4)	20	+1.2 (5.1)
Change from baseline to 18 months	11	+1.4 (3.0)	10	+2.7 (6.3)	15	+1.3 (8.3)
Change from baseline to 24 months	12	+2.0 (4.7)	9	+2.6 (6.6)	15	+1.1 (5.6)

Table 101 Short Form-36 (SF-36) physical component score results

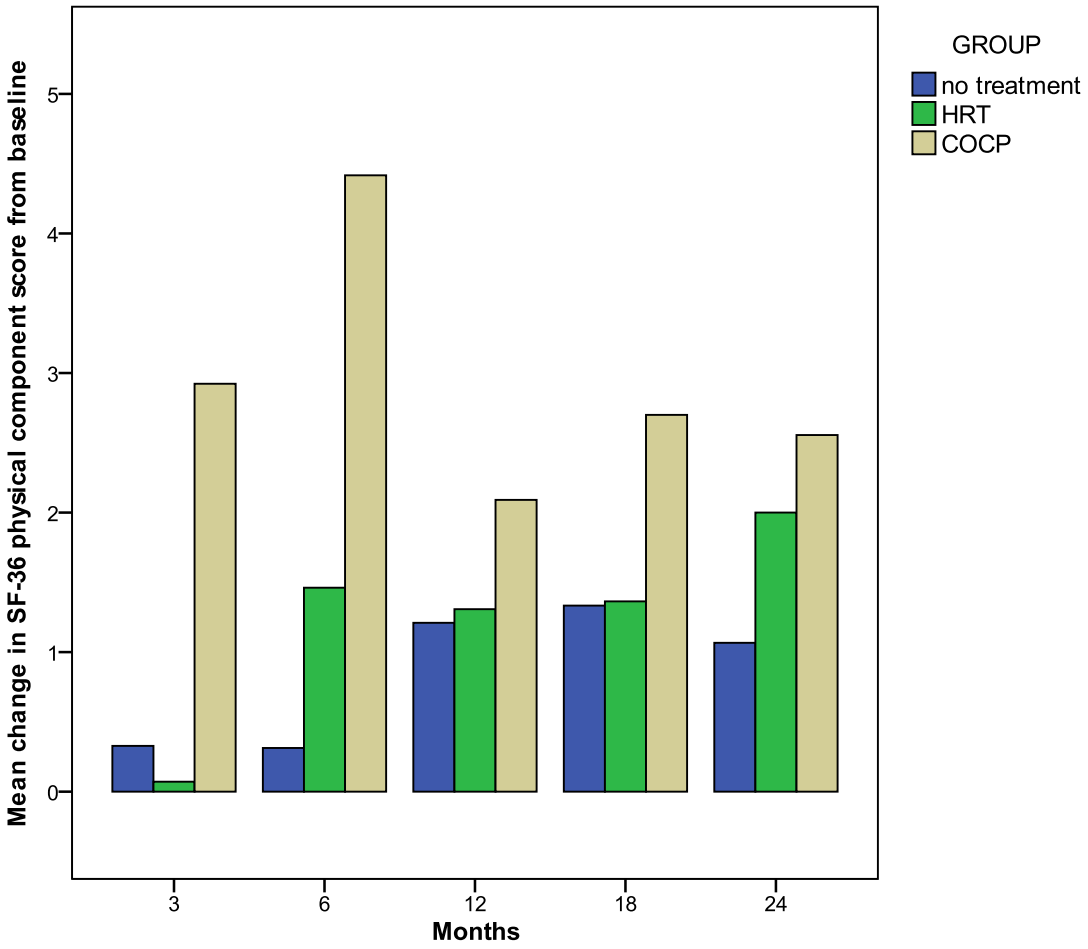


Figure 51 Changes from baseline in Short Form-36 (SF-36) physical component score over 24 months showing all available data at each time-point

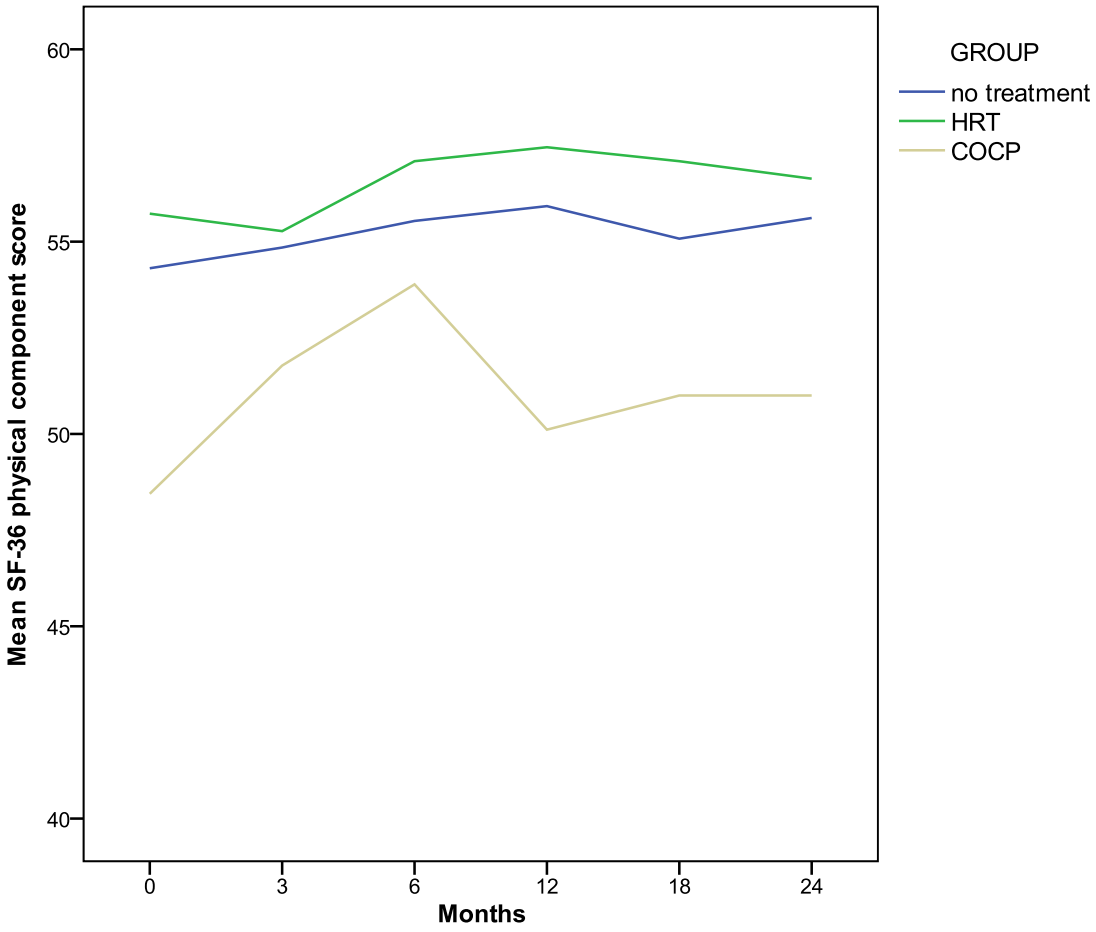


Figure 52 Changes in mean Short Form-36 (SF-36) physical component score over 24 months in participants with complete data collection

4.18.9.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean SF-36 physical component score score	95% confidence interval of the difference	p value
3	0.5	-2.6 to 3.6	0.756
6	1.5	-1.7 to 4.7	0.336
12	0.2	-4.0 to 4.4	0.929
18	-2.2	-6.2 to 1.9	0.273
24	-2.9	-7.2 to 1.4	0.169

Table 102 Comparison between HRT and COCP Short Form-36 (SF-36) summary physical health score results. Linear regression analysis was used to adjust for baseline score

4.18.9.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean SF-36 physical component score	95% confidence interval of the difference	p value
3	0.1	-3.2 to 3.4	0.948
6	1.7	-2.6 to 6.0	0.433
12	0.0	-3.4 to 3.5	0.978
18	1.0	-4.1 to 6.1	0.687
24	1.0	-3.2 to 5.2	0.625

Table 103 Comparison between HRT and no treatment Short Form-36 (SF-36) physical component score results. Linear regression analysis was used to adjust for baseline score

4.18.9.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean SF-36 physical component score	95% confidence interval of the difference	p value
3	0.8	-1.0 to 2.6	0.350
6	1.6	-0.7 to 3.9	0.167
12	0.4	-1.8 to 2.5	0.729
18	-0.1	-3.1 to 2.9	0.937
24	0.5	-2.3 to 3.3	0.708

Table 104 Comparison between COCP and no treatment Short Form-36 (SF-36) physical component score results. Linear regression analysis was used to adjust for baseline score

4.18.10 Mental Component Score

The mental component score increased in both the HRT and COCP groups, although it took longer (6 months) in the COCP group, as seen in some of the other questionnaires. There are no significant differences between the HRT and COCP groups. The changes in the HRT and no treatment groups are statistically significantly different at all time points except 24 months, when the p value is 0.070. Comparison between the COCP and no treatment groups did not show any differences.

	HRT		COCP		No treatment	
	Number of women	Mean (SD)	Number of women	Mean (SD)	Number of women	Mean (SD)
Baseline SF-36 mental component score	15	42.9 (13.9)	15	36.7 (16.3)	29	46.8 (10.0)
Change from baseline to 3 months	14	+6.9 (5.9)	13	+2.4 (9.0)	25	0.7 (8.0)
Change from baseline to 6 months	13	+6.5 (5.8)	12	+7.8 (8.9)	23	-0.5 (9.1)
Change from baseline to 12 months	13	+7.3 (8.0)	11	+6.5 (12.6)	20	-4.8 (10.7)
Change from baseline to 18 months	11	+5.3 (7.3)	10	+9.1 (18.5)	15	-3.8 (7.7)
Change from baseline to 24 months	12	+5.8 (6.4)	9	+9.9 (13.2)	15	0.8 (9.1)

Table 105 Short Form-36 (SF-36) mental component score results

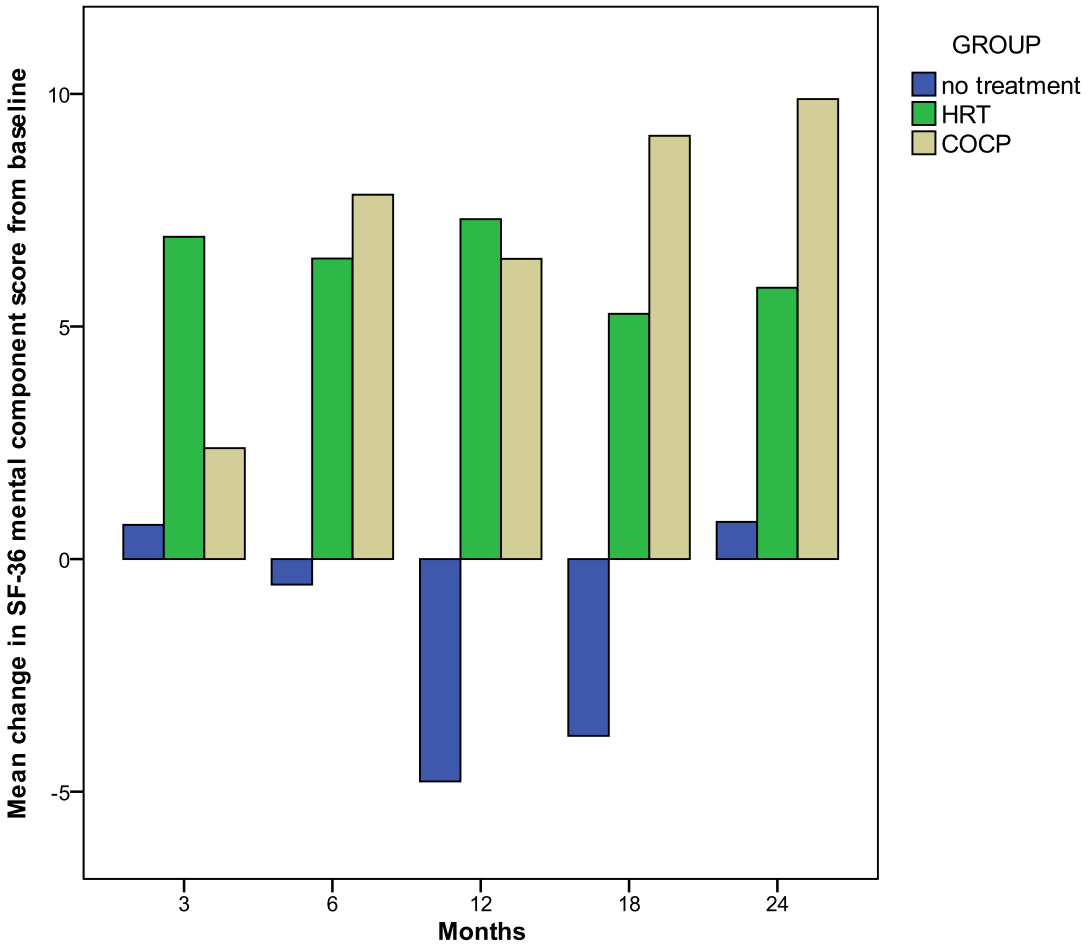


Figure 53 Changes from baseline in Short Form-36 (SF-36) mental component score over 24 months showing all available data at each time-point

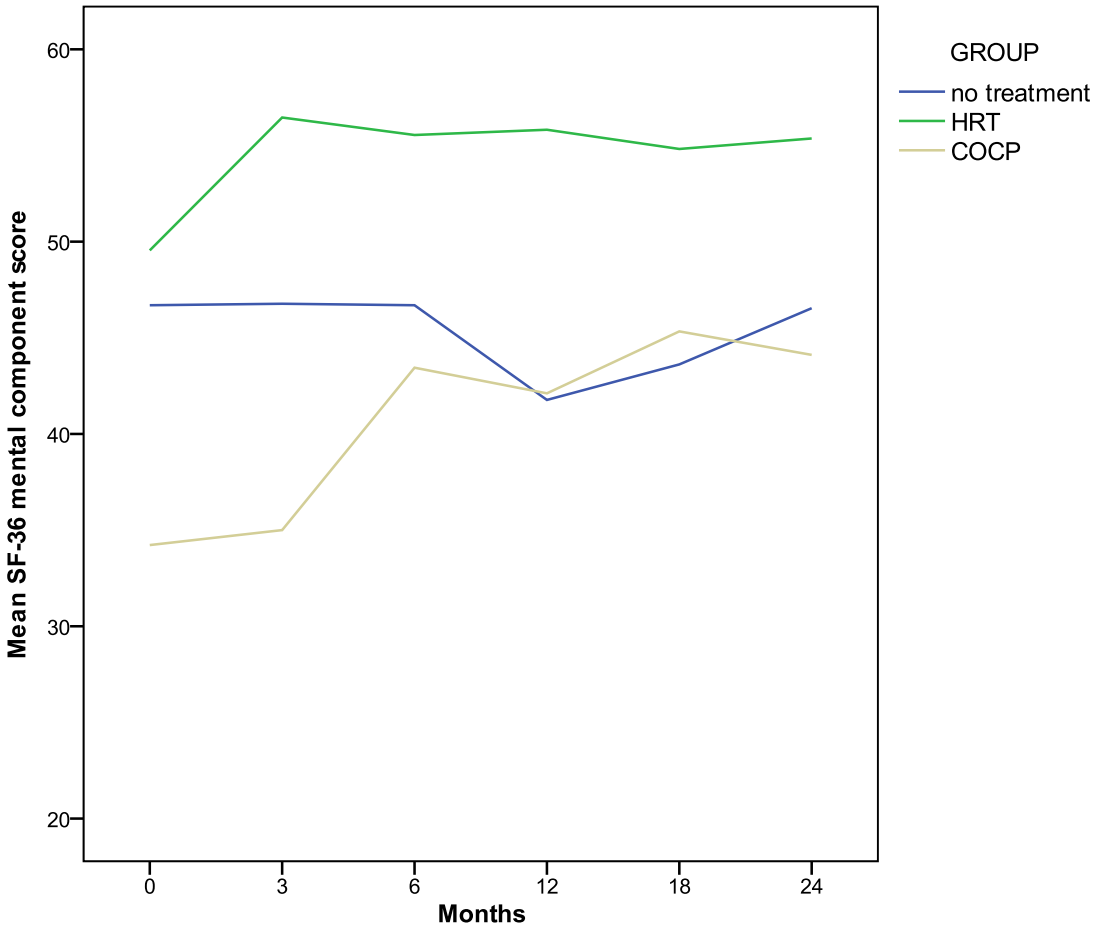


Figure 54 Changes in mean Short Form-36 (SF-36) mental component score over 24 months in participants with complete data collection

4.18.10.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean SF-36 mental component score	95% confidence interval of the difference	p value
3	-5.4	011.4 to 0.5	0.071
6	-1.2	-6.4 to 3.9	0.627
12	-4.0	-12.5 to 4.4	0.333
18	-6.5	-16.6 to 3.6	0.192
24	-1.0	-9.1 to 7.0	0.789

Table 106 Comparison between HRT and COCP Short Form-36 (SF-36) mental component score results. Linear regression analysis was used to adjust for baseline score

4.18.10.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean SF-36 mental component score	95% confidence interval of the difference	p value
3	5.7	0.9 to 10.5	0.021
6	6.2	0.7 to 11.8	0.029
12	11.5	4.6 to 18.5	0.002
18	10.1	4.2 to 16.1	0.003
24	5.5	-0.5 to 11.5	0.070

Table 107 Comparison between HRT and no treatment Short Form-36 (SF-36) mental component score results. Linear regression analysis was used to adjust for baseline score

4.18.10.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean SF-36 mental component score	95% confidence interval of the difference	p value
3	-0.1	-3.1 to 2.9	0.950
6	2.3	-1.1 to 5.6	0.176
12	4.1	-0.6 to 8.8	0.086
18	3.5	-1.6 to 8.6	0.168
24	2.2	-2.1 to 6.6	0.291

Table 108 Comparison between COCP and no treatment Short Form-36 (SF-36) mental component score results. Linear regression analysis was used to adjust for baseline score

4.18.11 Health Transition

Fig 55 shows that the health transition score in both the HRT and COCP groups improved over the course of the trial. By 6 months none of the participants in the treatment groups rated their health as worse than a year ago, although at 24 months one participant in the COCP group gave a score of somewhat worse than a year ago. More women in the HRT group than the COCP group rated their health as 'much better'. Improvements were also seen in the no treatment group, but this was largely due to women with lower scores withdrawing from the trial.

	Baseline	3 months	6 months	12 months	18 months	24 months
Much better (%)	0	21	23	23	18	33
Somewhat better (%)	27	29	31	23	36	25
About the same (%)	33	43	46	46	46	42
Somewhat worse (%)	20	0	0	8	0	0
Much worse (%)	20	7	0	0	0	0

Table 109 HRT group responses to the question 'Compared to one year ago, how would you rate your health in general now?'

	Baseline	3 months	6 months	12 months	18 months	24 months
Much better (%)	7	8	0	0	0	0
Somewhat better (%)	13	31	17	27	20	22
About the same (%)	47	46	75	64	60	44
Somewhat worse (%)	33	15	8	9	20	22
Much worse (%)	0	0	0	0	0	11

Table 110 COCP group responses to the question 'Compared to one year ago, how would you rate your health in general now?'

	Baseline	3 months	6 months	12 months	18 months	24 months
Much better (%)	17	12	9	10	13	0
Somewhat better (%)	17	16	17	20	20	20
About the same (%)	55	56	61	65	60	73
Somewhat worse (%)	7	12	13	5	7	7
Much worse (%)	3	4	0	0	0	0

Table 111 No treatment group responses to the question ‘Compared to one year ago, how would you rate your health in general now?’

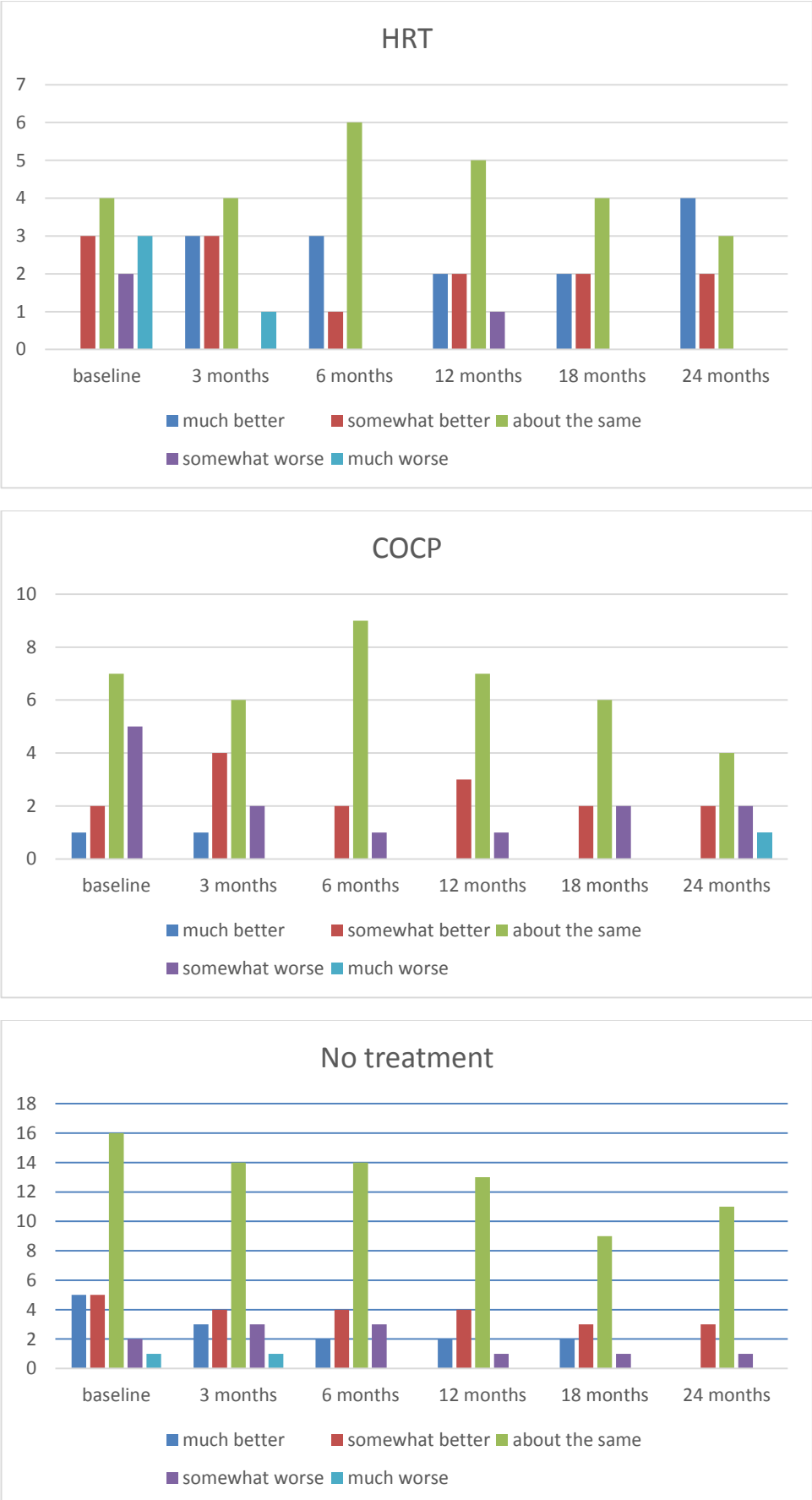


Figure 55 Short Form-36 Health Transition scores in the three groups across 24 months

4.19 Patient Health Questionnaire-9

The Patient Health Questionnaire-9 (PHQ-9) data were not normally distributed. Therefore data are shown as median (25%, 75%). Manipulation of the data was unable to produce a normal distribution and therefore the non-parametric Mann-Whitney U test was used to compare changes between groups. This does not provide estimation of difference in size effect between the groups and does not allow for adjustment for baseline score. This is particularly relevant because as shown in fig 57 the women who completed the trial in the COCP group had a higher baseline score than those in the HRT and no treatment groups. Fig 56 shows that there was a reduction in PHQ-9 score from baseline at all time-points in the HRT group and at all time-points except 3 months in the COCP group, but no changes were seen in the no treatment group.

	HRT		COCP		No treatment	
	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)	Number of women	Median (75%, 25%)
Baseline PHQ-9 score	15	7.0 (3.0, 17.0)	15	9.0 (2.0, 17.0)	29	5.0 (1.5, 9.5)
Change from baseline to 3 months	14	-3.0 (-7.8, -1.0)	13	0.0 (-4.0, 2.0)	25	0.0 (-4.0, 1.0)
Change from baseline to 6 months	13	-4.0 (-9.0, -0.5)	12	-5.0 (-7.0, 0.8)	23	0.0 (-3.0, 1.0)
Change from baseline to 12 months	13	-5.0 (-8.0, -1.0)	11	-3.0 (-6.0, 2.0)	20	0.0 (-0.75, 1.8)
Change from baseline to 18 months	11	-3.0 (-6.0, -1.0)	10	-3.5 (-6.3, 2.5)	15	0.0 (-2.0, 2.0)
Change from baseline to 24 months	12	-4.0 (-8.5, -0.3)	9	-2.0 (-10.0, - 2.0)	15	-1.0 (-5.0, 3.0)

Table 112 Patient Health Questionnaire-9 (PHQ-9) results

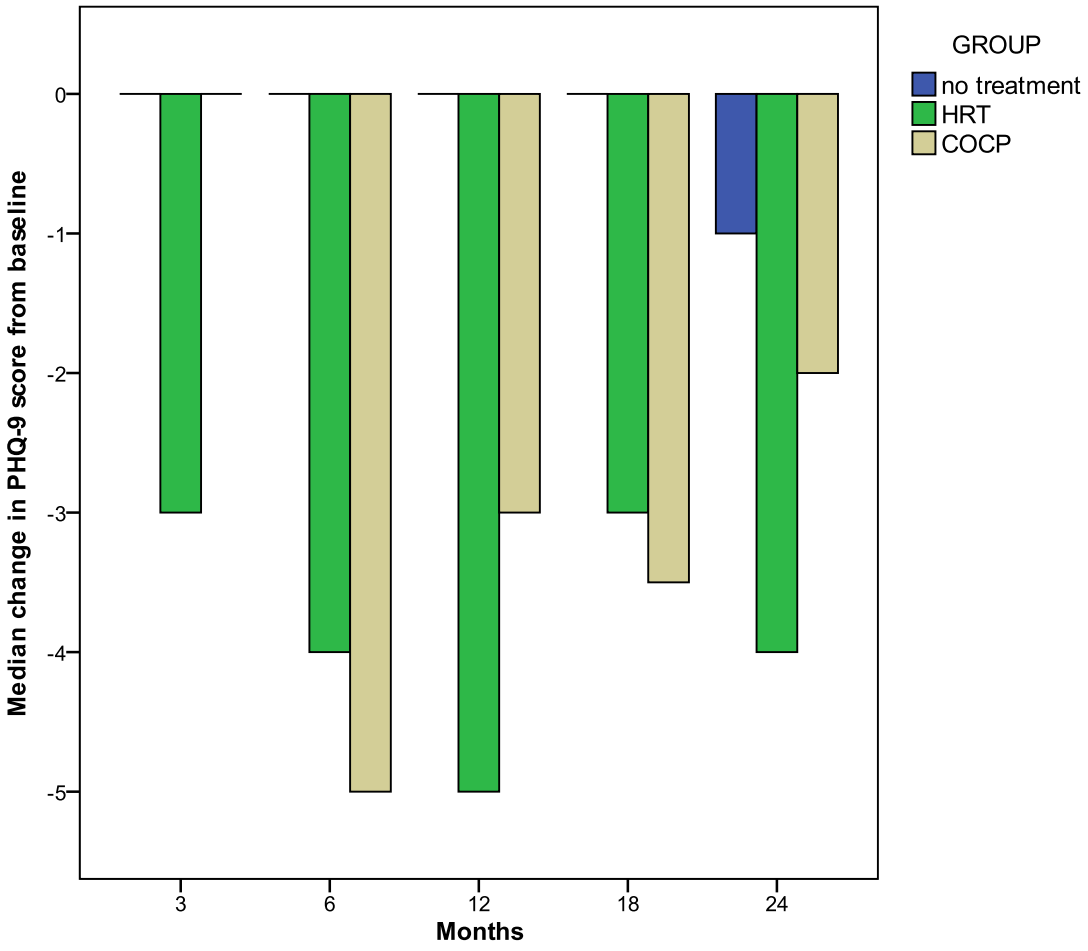


Figure 56 Changes from baseline in Patient Health Questionnaire-9 (PHQ-9) median score over 24 months showing all available data at each time-point

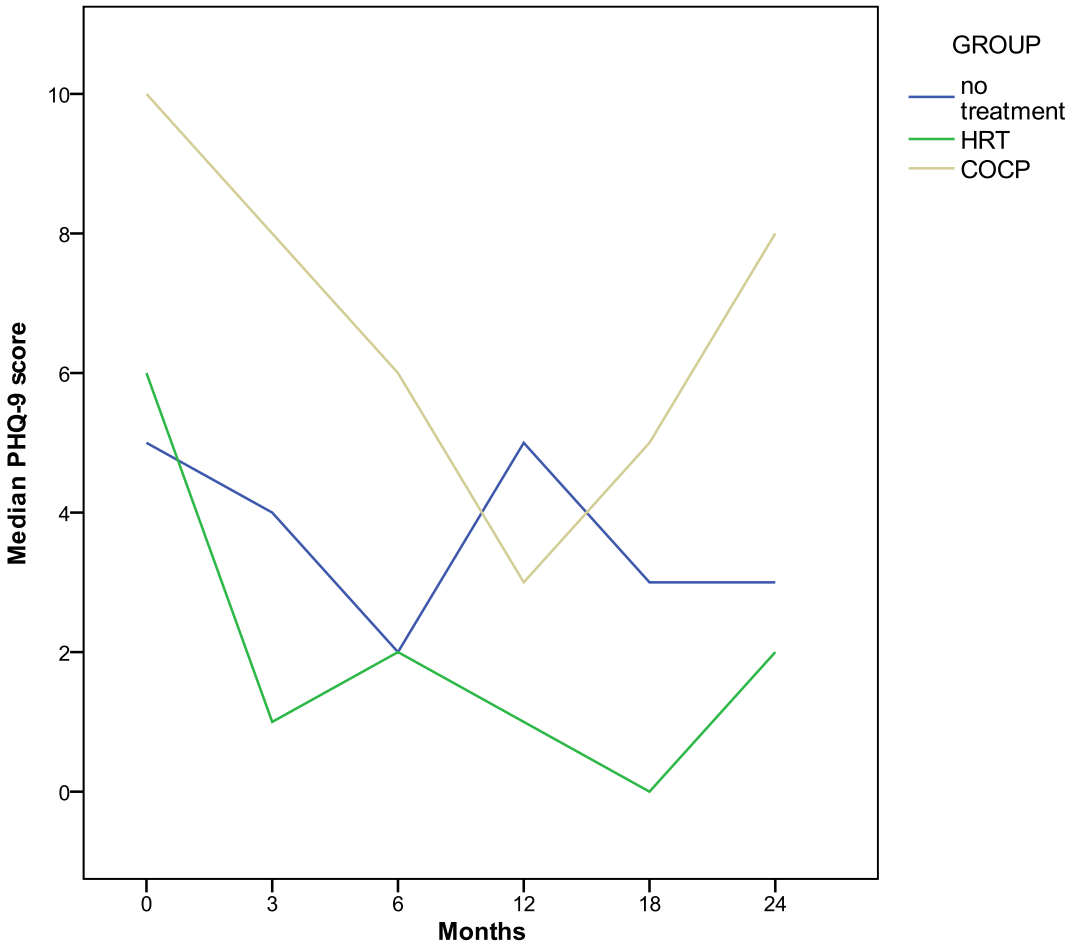


Figure 57 Changes in median Patient Health Questionnaire-9 (PHQ-9) score over 24 months in participants with complete data collection

Mann-Whitney U tests revealed no significant differences between the changes in score in the HRT and COCP groups at 3, 6, 12, 18 or 24 months (p values 0.064, 0.682, 0.309, 0.595 and 0.886 respectively). There were also no significant differences between the changes in the COCP and no treatment groups (p values 0.988, 0.090, 0.113, 0.191 and 0.134 at 3, 6, 12, 18 and 24 months). However, comparison between the HRT and no treatment groups revealed significant differences at all time points except 24 months (p values 0.024, 0.028, 0.001, 0.006 and 0.141 at 3, 6, 12, 18 and 24 months).

4.20 Blood pressure

Looking at all available data (fig 58 and 60), it appears that the COCP causes a slight rise in diastolic and systolic blood pressure as compared with HRT or no treatment. However, this is not seen on observing data solely from participants with complete data collection (figs 59 and 61). Significant differences were found on comparison between HRT and COCP at 3 months in both diastolic and systolic blood pressure. On comparison with the no treatment group, the only significant difference found was at 3 months between the COCP and no treatment groups. A lack of detection of further differences is likely to be due to the small numbers in this trial.

4.20.1 Systolic blood pressure

	HRT		COCP		No treatment	
	Number of women	Mean SBP (SD)	Number of women	Mean SBP (SD)	Number of women	Mean SBP (SD)
Baseline SBP	15	121.5 (11.0)	15	116.7 (8.2)	29	122.7 (13.9)
Change from baseline to 3 months	13	-1.0 (11.4)	13	+8.5 (11.4)	25	+0.8 (11.2)
Change from baseline to 6 months	13	-4.2 (13.5)	12	+2.3 (11.5)	23	-1.4 (12.9)
Change from baseline to 12 months	13	+1.7 (11.1)	11	+7.7 (8.7)	21	-1.9 (12.4)
Change from baseline to 18 months	11	-3.5 (15.7)	9	+2.1 (7.1)	15	-2.1 (10.9)
Change from baseline to 24 months	12	+3.1 (15.4)	9	+4.2 (9.4)	15	+0.3 (9.9)

Table 113 Systolic blood pressure (SBP) results (mmHg)

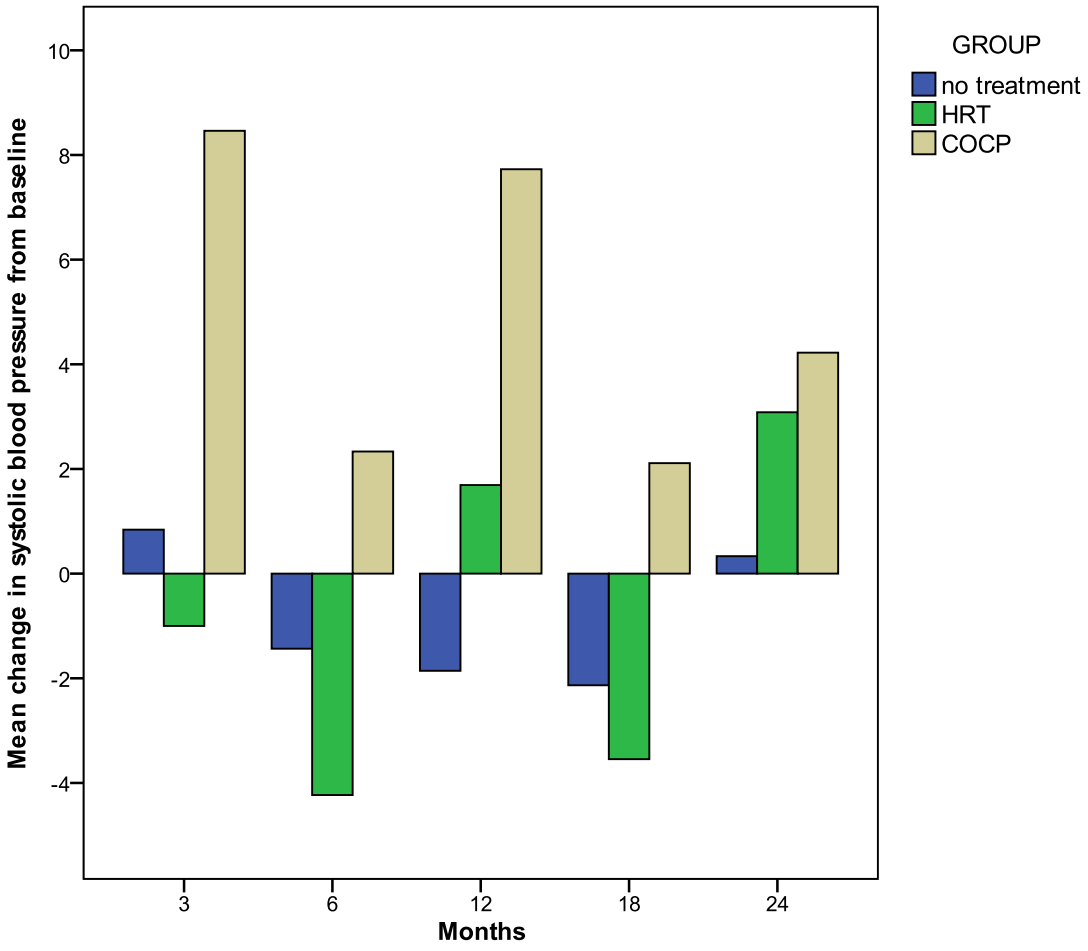


Figure 58 Changes from baseline in mean systolic blood pressure (mmHg) over 24 months showing all available data at each time-point

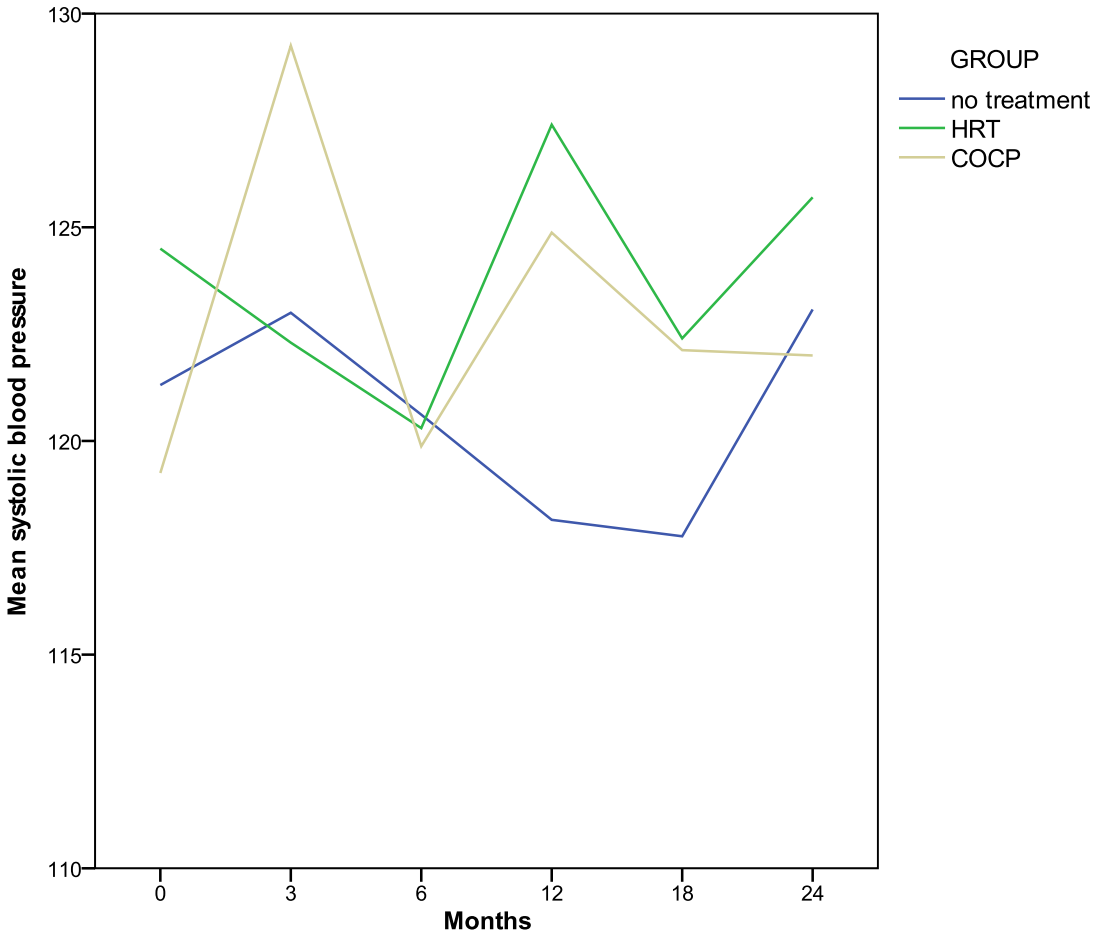


Figure 59 Changes in mean systolic blood pressure (mmHg) over 24 months in participants with complete data collection

4.20.1.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean systolic blood pressure	95% confidence interval of the difference	p value
3	10.2	0.6 to 19.8	0.038
6	5.2	-5.4 to 15.8	0.317
12	5.2	-3.6 to 14.1	0.231
18	4.1	-8.1 to 16.3	0.492
24	0.3	-12.2 to 12.9	0.959

Table 114 Comparison between HRT and COCP systolic blood pressure (mmHg) results. Linear regression analysis was used to adjust for baseline score.

4.20.1.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean systolic blood pressure	95% confidence interval of the difference	p value
3	-2.0	-9.2 to 5.1	0.573
6	-3.0	-11.3 to 5.4	0.475
12	3.1	-4.6 to 10.7	0.423
18	0.3	-9.1 to 9.7	0.949
24	2.8	-7.0 to 12.6	0.562

Table 115 Comparison between HRT and no treatment systolic blood pressure (mmHg) results. Linear regression analysis was used to adjust for baseline score.

4.20.1.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean systolic blood pressure	95% confidence interval of the difference	p value
3	3.1	-0.7 to 6.9	0.108
6	0.6	-3.2 to 4.5	0.745
12	3.4	-0.2 to 6.9	0.062
18	1.8	-1.5 to 5.2	0.272
24	1.4	-2.5 to 5.3	0.465

Table 116 Comparison between COCP and no treatment systolic blood pressure (mmHg) results. Linear regression analysis was used to adjust for baseline score.

4.20.2 Diastolic blood pressure

	HRT		COCP		No treatment	
	Number of women	Mean SBP (SD)	Number of women	Mean SBP (SD)	Number of women	Mean SBP (SD)
Baseline DBP	15	77.3 (11.2)	15	76.9 (12.5)	29	80.5 (11.7)
Change from baseline to 3 months	13	-3.6 (5.1)	13	+3.1 (5.0)	25	-4.5 (9.8)
Change from baseline to 6 months	13	-2.5 (9.9)	12	-1.8 (13.8)	23	-3.7 (10.1)
Change from baseline to 12 months	13	+0.8 (7.8)	11	+1.1 (15.4)	21	-4.0 (9.0)
Change from baseline to 18 months	11	-0.3 (9.5)	9	+3.7 (7.8)	15	-3.0 (11.3)
Change from baseline to 24 months	12	+0.7 (6.4)	9	+4.8 (10.1)	15	-2.7 (11.1)

Table 117 Diastolic blood pressure (DBP) results (mmHg)

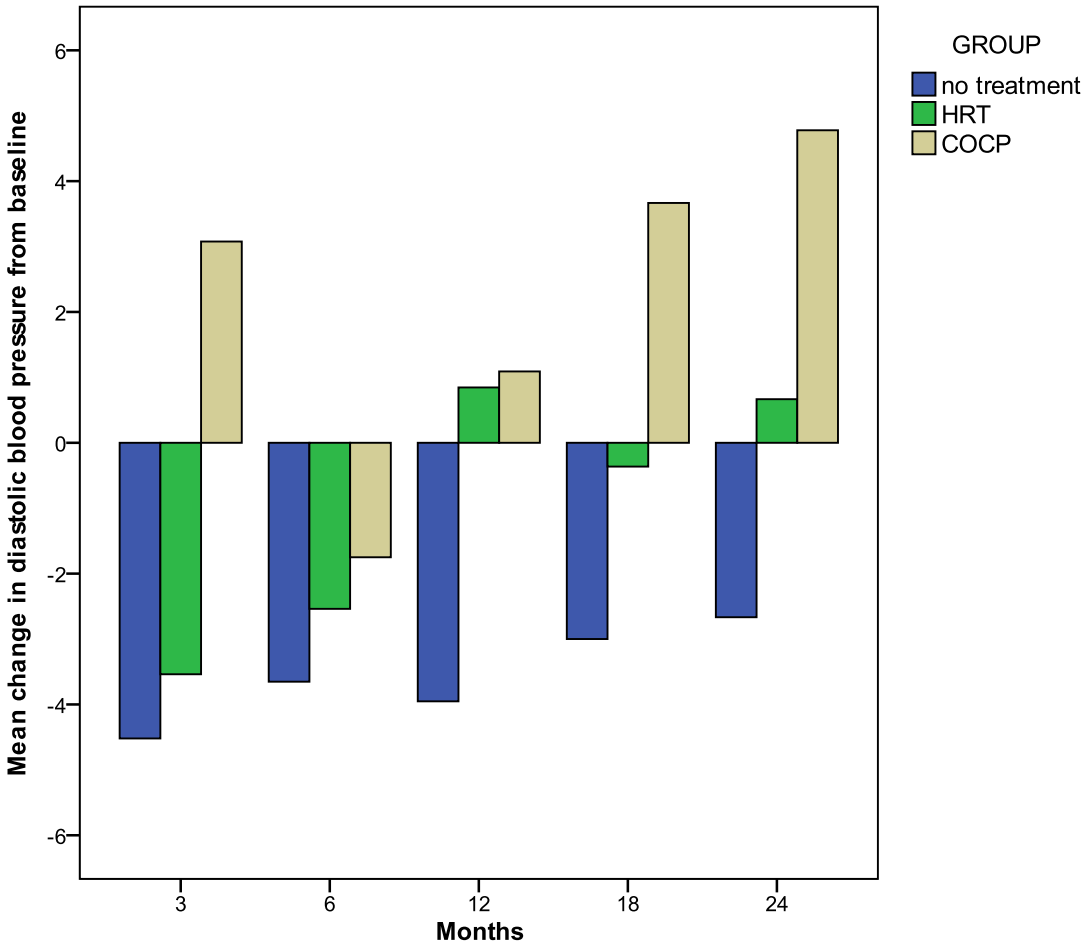


Figure 60 Changes from baseline in mean diastolic blood pressure (mmHg) over 24 months showing all available data at each time-point

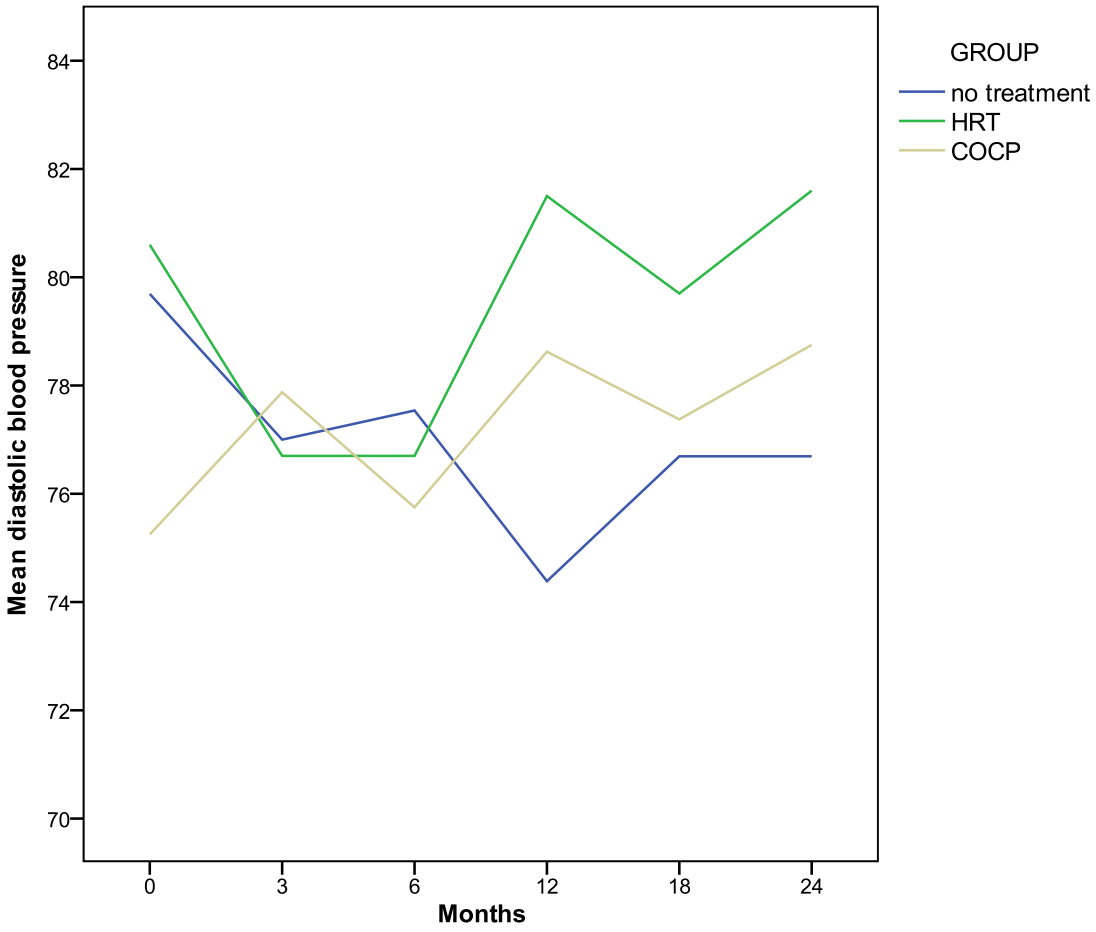


Figure 61 Changes in mean diastolic blood pressure (mmHg) over 24 months in participants with complete data collection

4.20.2.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean diastolic blood pressure	95% confidence interval of the difference	p value
3	6.0	2.0 to 9.9	0.005
6	1.5	-6.1 to 9.0	0.694
12	0.5	-7.2 to 8.3	0.888
18	2.7	-5.2 to 10.6	0.483
24	4.2	-3.7 to 12.0	0.281

Table 118 Comparison between HRT and COCP diastolic blood pressure (mmHg) results. Linear regression analysis was used to adjust for baseline score.

4.20.2.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean diastolic blood pressure	95% confidence interval of the difference	p value
3	0.3	-4.3 to 4.8	0.913
6	-0.4	-6.3 to 5.5	0.884
12	3.8	-1.6 to 9.2	0.159
18	2.6	-4.6 to 9.8	0.458
24	3.1	-3.9 to 10.0	0.372

Table 119 Comparison between HRT and no treatment diastolic blood pressure (mmHg) results. Linear regression analysis was used to adjust for baseline score.

4.20.2.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean diastolic blood pressure	95% confidence interval of the difference	p value
3	2.9	0.2 to 5.5	0.035
6	0.4	-3.1 to 4.0	0.814
12	1.9	-1.7 to 5.5	0.287
18	2.6	-1.5 to 6.7	0.202
24	2.8	-1.4 to 7.1	0.181

Table 120 Comparison between COCP and no treatment diastolic blood pressure (mmHg) results. Linear regression analysis was used to adjust for baseline score.

4.21 Body mass index

Both the graph showing changes from baseline using all available data at each time-point (fig 62) and the graph displaying results from participants with complete data collection (fig 63) show that in the COCP and no treatment groups the body mass index (BMI) had an upwards trend. This was more marked in the COCP group. In the HRT group the BMI remained relatively constant. However, when compared formally there were no statistically significant differences between any of the groups.

	HRT		COCP		No treatment	
	Number of women	Mean BMI (SD)	Number of women	Mean BMI (SD)	Number of women	Mean BMI (SD)
Baseline BMI	15	25.1 (3.6)	15	23.9 (4.6)	29	24.2 (3.2)
Change from baseline to 3 months	12	-0.04 (0.44)	13	-0.03 (0.86)	24	0.25 (0.78)
Change from baseline to 6 months	13	0.51 (0.89)	12	0.49 (0.91)	23	0.43 (1.11)
Change from baseline to 12 months	12	0.51 (1.05)	11	0.77 (0.98)	20	0.35 (0.92)
Change from baseline to 18 months	11	0.05 (0.92)	10	0.61 (0.87)	15	0.70 (1.58)
Change from baseline to 24 months	12	0.23 (1.28)	9	0.90 (1.13)	15	0.42 (1.41)

Table 121 Body mass index (BMI) results (kg/m^2)

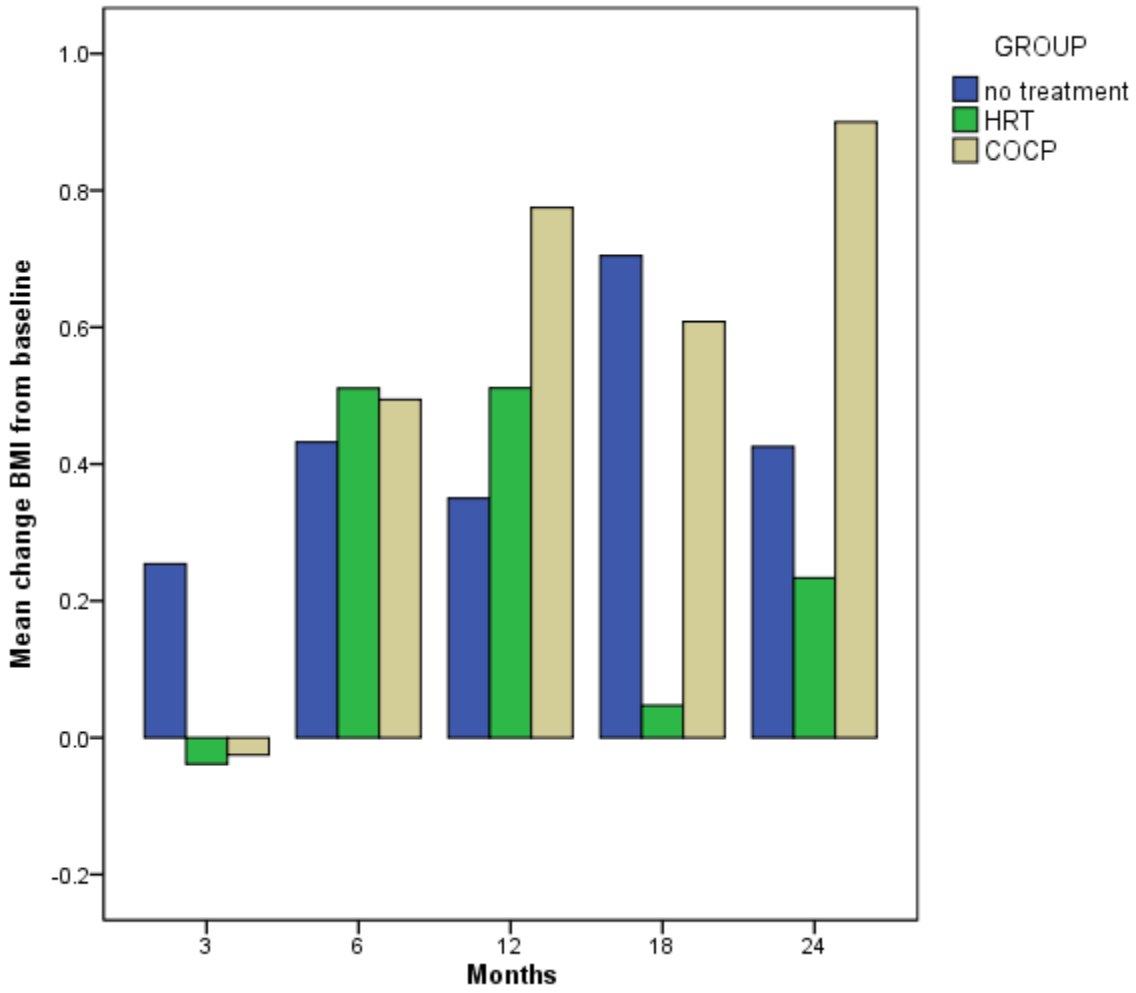


Figure 62 Changes from baseline in body mass index (BMI) (kg/m^2) mean score over 24 months showing all available data at each time-point

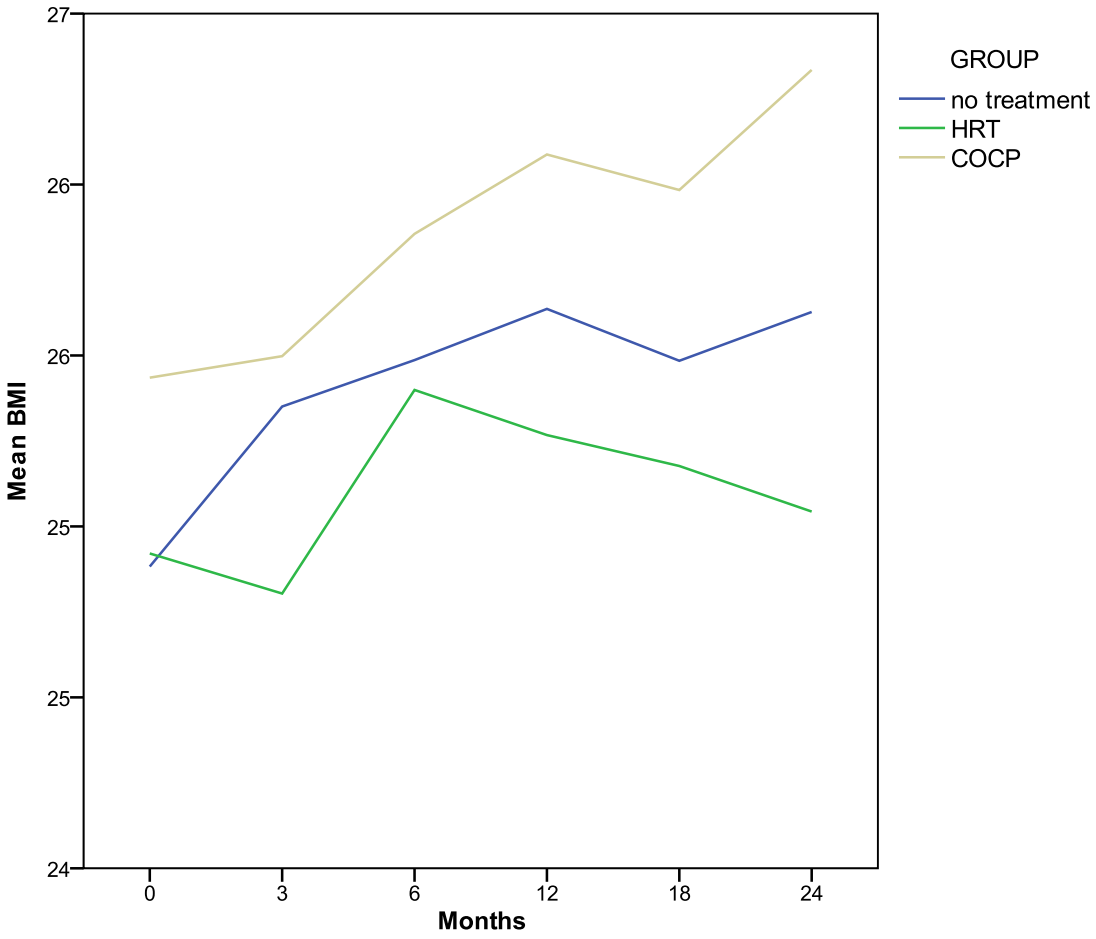


Figure 63 Changes in mean body mass index (BMI) (kg/m^2) score over 24 months in participants with complete data collection

4.21.1 Comparison between HRT and COCP groups

Months	COCP minus HRT mean BMI (kg/m^2)	95% confidence interval of the difference	p value
3	-0.0	-0.6 to 0.5	0.899
6	-0.0	-0.8 to 0.7	0.929
12	0.3	-0.6 to 1.2	0.555
18	0.6	-0.3 to 1.4	0.180
24	0.7	-0.4 to 1.8	0.216

Table 122 Comparison between HRT and COCP body mass index (BMI) (kg/m^2) results. Linear regression analysis was used to adjust for baseline score.

4.21.2 Comparison between HRT and no treatment groups

Months	HRT minus no treatment mean BMI (kg/m^2)	95% confidence interval of the difference	p value
3	-0.3	-0.8 to 0.2	0.270
6	0.1	-0.7 to 0.8	0.841
12	0.2	-0.6 to 0.9	0.653
18	-0.6	-1.7 to 0.5	0.254
24	-0.2	-1.2 to 0.9	0.765

Table 123 Comparison between HRT and no treatment body mass index (BMI) (kg/m^2) results. Linear regression analysis was used to adjust for baseline score.

4.21.3 Comparison between COCP and no treatment groups

Months	COCP minus no treatment mean BMI (kg/m^2)	95% confidence interval of the difference	p value
3	-0.1	-0.4 to 0.1	0.363
6	0.0	-0.4 to 0.4	0.886
12	0.2	-0.2 to 0.6	0.257
18	-0.0	-0.6 to 0.5	0.895
24	0.3	-0.3 to 0.8	0.368

Table 124 Comparison between COCP and no treatment body mass index (BMI) (kg/m^2) results. Linear regression analysis was used to adjust for baseline score.

Chapter 5 Discussion

The main initial obstacle to progression of the trial was recruitment, which was much slower than anticipated. Recruitment to the treatment group was particularly difficult as there were few recently diagnosed women who wanted to take treatment but did not mind if they took HRT or the COCP. Many did not want to jeopardise the small but real possibility of spontaneous pregnancy with the chance of being randomised to COCP, but would have participated if the study had been comparing two different HRT preparations. There were also women who were already settled on treatment and did not want to complete a two month washout period prior to randomisation. We anticipated that there would be fewer women wanting to join the no treatment group, but in fact women who had already decided they did not wish to take treatment welcomed the opportunity to be closely monitored and recruitment to this group was completed first. Women choosing no treatment were fully counselled on the benefits of oestrogen replacement both before taking part and throughout the trial. Despite spending more time discussing this than would normally be available in clinic, only a few changed their minds and opted for treatment. There will always be women who decline treatment for various reasons and this trial provides valuable data on what these women can expect, especially regarding bone density and symptoms. It is important to remember that these women continued to choose to decline treatment at each time-point and that there were several who withdrew because of symptoms or a drop in bone density causing them to opt for treatment. Recruitment was closed after two years due to time constraints and at this point there were 15 women (rather than the planned 22) in each treatment group. The three year from diagnosis cut-off for recruitment into the study was chosen because bone density changes the most in the early years following diagnosis. However in retrospect, particularly for the treatment groups, it may have been reasonable to expand this to five years from diagnosis. The use of varied preparations of COCP and HRT may also have improved recruitment and also retention in the study as some women did not want to take a particular brand or wanted the option to change if it did not suit them, as is common in clinical practice. However this would then have made interpretation of the results more difficult, particularly with the small numbers in the trial.

The main limitations of this trial were the small sample size and the high drop-out rate, which unfortunately occurred to a greater extent in the COCP than the HRT group. One participant in each of these groups withdrew due to symptoms of depression and one in each group because of side effects. The excess drop-out rate in the COCP group was due to loss of follow up. Every effort was made to determine other possible reasons for withdrawal but none could be obtained. There is therefore a higher level of missing data in the COCP group. One method to accommodate for missing data would have been multiple imputation, which is the process of filling in missing data using probable values gained from all other information (164). In the absence of expert dedicated statistical assistance this was not feasible in this trial but it would need to be a consideration in planning further trials in this area. In some fields women with complete data collection in the COCP group appear to have different baseline values compared with women with complete data collection in the HRT group. This is particularly striking in the menopausal symptoms and depression questionnaires and the high sensitivity C-reactive

protein but not seen in the lumbar spine bone density or lipid results. As the groups were randomised at baseline, it can be assumed that values were comparable at this point and this is confirmed by the baseline data tables. Differences in baseline values in women with complete data collection are therefore due to differences in those who withdrew. This point has been illustrated by displaying graphs throughout the results showing not only changes from baseline using all available data at each time-point but also showing the absolute values for each variable amongst participants with complete data collection. The possibility that women in the HRT group with poor baseline symptom scores were lost to follow up because of a lack of improvement of their symptoms cannot be excluded. However, this would not explain why women in a similar position in the COCP group did not withdraw. It seems likely that baseline differences in women who withdrew from each group are due to chance. These baseline differences are compensated to some extent by using adjustment for baseline score in statistical analysis where possible (for variables with normal distributions). Although the drop-out rate is high, it is worth noting that in the only other similar trial to this, there was a drop-out rate of 47% (83), compared with 30% in the treatment arms of this trial. Women received continuity of care and flexibility with appointments and it is difficult to see how loss of follow up where no reason was given could have been improved. This will need to be considered in planning future trials in this groups of women. The compliance with medication calculated from returned pills was very high (over 95%). Whilst we did not perform independent markers of compliance with medication such as FSH levels, unscheduled bleeding was considered to be an adverse event and was not reported in either treatment group, suggesting good compliance.

Another limitation is that the no treatment group was not randomised. The main comparisons of value are those between the HRT and COCP groups. However, comparisons between each of these groups and the no treatment group have been included because they illustrate differences between women who choose (and continue to choose) to take no treatment and those who choose to take treatment and are randomised to a particular group. In some parameters, there can be no doubt that treatment is highly beneficial. For example in the bone density results, the bone density at the hip is maintained in both the HRT and COCP groups and falls at every time-point in the no treatment group. Statistical comparison reveals highly significant differences at all time-points between the no treatment group and each of the treatment groups. At the spine, HRT is again clearly better than no treatment, but although COCP shows a trend towards benefit, the results are not statistically significantly different. A larger sample size would be needed to detect a difference.

Of the few trials in the area of POF, most use a cut-off age of under 40 rather than under 45. An age of 44 or less was used in this trial because the primary outcome was bone density and this, along with many of the other outcomes in this study, continues to be a concern for women with ovarian failure in their early 40s. Using an age limit of 44 made recruitment possible; if we had used an upper limit of 39 only 27 women would have been recruited. The median age in this study was 40.5. This age group of women have busy lives and frequently needed to re-arrange their appointments because of work or family life. Those who withdrew were not followed up and so an intention to treat analysis was not possible.

Baseline data in this trial reveal some interesting results. For example, looking at data from all groups, 60% of women at baseline report reduced libido, 58% a lack of energy and 58% irritability. These are symptoms which may not be covered in a routine consultation, especially in the time constraints of general practice, but have the potential to have a large impact on quality of life. Menopausal symptom depression scores were higher and mental health scores lower in the women who opted for treatment rather than no treatment. Of note, there were no significant differences in bone density in women choosing no treatment or treatment and the mean baseline Z scores were -0.6, -0.5 and -0.6 in the HRT, COCP and no treatment groups respectively. This is higher than found in most cross-sectional studies investigating BMD in POF and is presumably due to the relatively short time since diagnosis.

In the bone density results, the most significant findings were at the lumbar spine, which is recognised to be the best site for monitoring response to treatment (64). HRT increased bone density at the lumbar spine whereas COCP only maintained it and this was significant at 12 and 24 months. Importantly, in participants with complete data collection the baseline lumbar spine BMD was similar in the HRT and COCP groups and so the differences cannot be explained by the HRT group having a significantly lower initial bone density. However, the bone marker results did not show any significant differences between the HRT and COCP groups. The lack of a significant difference between the treatment groups in the bone markers is likely to be due to the high variability of these markers. The only other prospective trial comparing HRT and COCP in POF (Crofton et al) investigated bone density and bone turnover markers in a crossover trial with 12 months each of HRT and COCP; differences were found in markers of bone formation (ALP and P1NP) in favour of HRT but no significant differences were found in bone density or CTX (83). This discrepancy with our results could be explained by the relatively higher doses of oestrogen in the HRT used in Crofton et al's study (transdermal oestradiol 100mcg for 1 week then 150mg for weeks 2-4). A lack of difference between COCP and HRT in the bone markers in our trial means that the differences in bone density between these groups need to be interpreted with caution, especially in light of the high drop-out rate in the COCP group. However, in the absence of any other data from clinical trials, it seems reasonable to inform women that HRT appears to have a better effect on bone density than the COCP. A recent Cochrane review on contraceptives and bone density did not find any effect of combined oral contraceptives on bone density in healthy adults but acknowledged that bone density may be affected by COCP use in adolescent and young women (165). It is important to remember that some women with POF will be starting oestrogen replacement before they have reached their peak bone mass and taking it for many years and whilst the COCP may be a more natural choice for this age group, it is probably not the best oestrogen replacement for their bone development. Further research in this group of women in particular is required.

There are several possible reasons for the significant difference between HRT and COCP in lumbar spine bone density. The COCP was used in its traditional way, with 3 weeks of pills followed by a hormone free week. It is possible that this oestrogen-free week affected bone mass acquisition. The types of oestrogen are also different and oestradiol may have a more favourable effect on bone than ethinyloestradiol. The preparations both contain levonorgestrel

but at different doses (75mcg in Nuvelle and 150mcg in the COCP), and with Nuvelle progesterone is only taken for 12 days a month rather than 21. Although oestrogens are usually considered to be the main factor in HRT affecting bone density, levonorgestrel is an androgenic progesterone and may affect the bone in high concentrations (166). There is a theory is that in lower doses levonorgestrel has a direct stimulatory effect on osteoblasts via progesterone and androgen receptors and can help to increase bone mass, but at higher doses some can also bind to glucocorticoid receptors and inhibit osteoblasts in a dose-dependent manner.

In the no treatment group, there was a reduction in bone density at all sites by 24 months and significant differences in bone markers when compared with both HRT and COCP at each time point. From these results it is reasonable to conclude that women who choose not to take any form of hormone replacement will suffer adverse effects on their bone health compared with those who opt for either HRT or the COCP. Women should be urged to take some form of oestrogen replacement rather than none.

It was very disappointing that all but one of the AMH levels was undetectable. Previous cross-sectional studies investigating AMH in POF have shown detectable levels in up to 23% (24, 133). Even accounting for an older age group in this trial, it is surprising that there are not more detectable levels. Although the product literature advises that samples are stable when frozen, concerns have recently been raised about the stability of storage of samples which are analysed using the AMH Gen II assay (167, 168). Samples analysed in this study used the recommended pre-mix protocol to eliminate complement interference which can cause falsely low AMH levels, which should have eliminated this problem.

Inhibin B has been reported as an inferior marker of ovarian failure (120), although it has been noted to relate to the presence of ovarian follicles in POF (134). It is interesting that out of the 17 participants who had at least one detectable level, 11 had a detectable level on another occasion. Inhibin B may warrant further investigation in POF in the detection of intermittent ovarian function. Only one participant had a detectable inhibin B level at every visit. There were more detectable levels at baseline than at 24 months (even allowing for drop-outs), which suggests there is more ovarian activity close to the time of diagnosis, as has been previously described (3).

Although there were no statistically significant differences between any of the groups on lipid profile, the trend of reduction of LDL with HRT was seen in another study in POF patients (98). This study also reported an increase in TG with HRT which was not seen in our study. The trends we found in LDL and HDL were in favour of HRT compared with COCP. This is as expected with the oestrogens and progestogen in the preparations used (169, 170). High-sensitivity CRP was increased in the COCP group, but this is difficult to interpret because this group had a lower level at baseline and the non-parametric distribution of the data (which could not be transformed to a normal distribution) meant that regression analysis with adjustment for baseline values was not possible.

The Modified Greene Climacteric Scale (MGCS) results showed rapid reductions in symptom scores following initiation of HRT, as seen at the time of the normal menopause (53). Reductions in symptom score were also seen with the COCP, but maximal effects were seen at 6 months rather than 3 months with HRT; the exception to this was vasomotor symptoms which were controlled equally quickly. This delayed effect in the COCP group was also observed in the Menopause Symptoms Treatment Satisfaction Questionnaire results and the Mental Component Score of the Short Form-36. MGCS scores at 24 months indicate better symptom control with HRT compared with COCP in all domains except vasomotor and sexual function, which as a single question is unlikely to be sensitive enough to ascertain a difference. Baseline MGCS scores were similar to those observed in general menopause clinics. The Menopause Symptoms Treatment Satisfaction Questionnaire results are difficult to interpret in view of the discrepancy between baseline scores in those who withdrew from each of the treatment groups. As there is no baseline score adjustment for this cannot be performed. Despite this, there appears to be a higher satisfaction with treatment at each time point in the HRT group, and this is highly significant at 3 months (p 0.003) when there was a good amount of data collection, confirming the findings in the MGCS of a delay in symptom control with the COCP.

The results from the Brief Profile of Female Sexual Function questionnaire are very interesting, especially in view of sexual function being identified as a matter of high concern to women with POF (55). The baseline scores of 18 in both the treatment groups and 20 in the no treatment group show high levels of sexual dysfunction (a score of 20 or less indicates hypoactive sexual desire disorder). From a similar baseline score in the HRT and COCP groups, even amongst women who completed the trial, the HRT group's score increased whereas that of the COCP group remained stable. However these differences did not reach statistical significance. Sexual function in the COCP and no treatment groups remained similar to baseline levels throughout the trial. In the general population it is recognised the COCP increases sex hormone binding globulin which in turn reduces the free testosterone level (171) and this has been found to be associated with the presence of hypoactive sexual desire disorder (172). A 2009 internet-based survey on sexual dysfunction including 2527 peri- and post-menopausal women reported that over one third of those who had tried HRT found that it improved symptoms of sexual dysfunction (173). Although the results in our study did not reach statistical significance, in view of other evidence, it seems reasonable to suggest a potential benefit of HRT over the COCP for women with POF who have a particular issue with sexual dysfunction.

The Short Form-36 quality of life questionnaire did not show any consistent differences between the groups in most domains. In the vitality domain, there was a non-significant improvement in score in both of the treatment groups. The statistical comparison between the HRT and no treatment groups in this domain was significant at all but one time-point, but no differences were demonstrated between the COCP and no treatment groups. In the Role Emotional domain, the COCP was significantly better than HRT at 24 months, although not at other time-points, and the COCP was better than no treatment at every time-point except 3 months. The mental health score improved in both the HRT and COCP groups, with no significant differences between the two. When compared with the no treatment group only HRT vs no treatment was

significant. In response to the health transition question, at 24 months in the HRT group 33% rated their health as much better than a year ago; 25% said it was somewhat better; 42% reported it was about the same; 0% somewhat worse and 0% much worse. By comparison, in the COCP group the figures were 0% (much better); 22% (somewhat better); 44% (about the same); 22% somewhat worse and 11% much worse. The baseline scores in the COCP group were lower than those in the HRT group and the scores in both treatment groups improved over the course of the trial.

The Patient Health Questionnaire-9 assessment of depression shows reduction in scores in the HRT and COCP groups but not in the no treatment group. Significant differences were found between the HRT and no treatment groups at every time-point except 24 months. Unfortunately as the data were not normally distributed there could be no adjustment for baseline score, which was higher in the COCP group in women who completed the trial.

The COCP group had a slight rise in both systolic and diastolic blood pressure, evident from 3 months, although it was only significantly different from the HRT group at 3 months. This was not seen in the HRT or no treatment groups, whose values remained relatively stable throughout the trial. A similar increase in blood pressure with COCP was seen in a crossover trial of women with POF taking COCP and HRT for a year each (101); 18 women completed both regimens and a mean difference of systolic blood pressure of 7.3mmHg and diastolic blood pressure of 7.4mmHg were found following 12 months treatment. These are similar to the numbers obtained in our study at 3 months and it is likely that significant differences were not found later on due to reduced numbers of women continuing in the trial. Body mass index had an upward trend in the COCP and no treatment groups and remained stable in the HRT group, but none of the groups had significantly different changes on statistical analysis. A larger sample size would be needed to evaluate this further.

Chapter 6 Conclusion

Although this trial has limitations of small sample size and high drop-out rate in the COCP group, it makes a significant contribution to the very small body of evidence concerning the treatment of premature ovarian failure. There are very few randomised trials, or even cohort studies, in this area. It is important when interpreting the results to remember that the no treatment group was self-selected and actively chose to continue not to take treatment at each time-point.

Comparison between the HRT and COCP groups revealed a significant difference in lumbar spine BMD at 24 months in favour of HRT (0.038 g/cm^2; 95% CI 0.002 to 0.073; p 0.040), which was the primary outcome for this study. However there were no significant differences between the HRT and COCP groups' bone marker results.

We found that women who chose not to take any hormonal treatment in POF were as expected actively losing bone. This group experienced a drop in bone density at the lumbar spine, total hip and femoral neck. Differences are highly significant between HRT and no treatment groups at the lumbar spine and between both COCP and HRT versus no treatment at the total hip. These findings were supported by results from bone marker analysis.

HRT performed favourably compared with the COCP in lipid profile, sexual function, blood pressure and depression scores, although most differences were not statistically significant. Larger sample sizes are needed to confirm and quantify differences and results from this trial could be used to help calculate sample sizes for future trials. HRT also appeared to provide a more rapid control of menopausal symptoms, with improvements from 3 rather than 6 months.

In summary, it appears that HRT has a more favourable effect than COCP on bone density and several other parameters in POF. However, it is also clear that choosing to take no treatment has a poor effect on bone health and symptoms and it is preferable to take any form of hormone treatment rather than none. Some women will not find HRT an acceptable medication to take and it would be better for these women to take the COCP than nothing at all.

It is also highly likely that different doses and routes of oestrogen replacement may have different effects on bone density and metabolism and this is an area that deserves further investigation. There is currently one trial underway which is evaluating the effects of micronised progesterone and medroxyprogesterone acetate, both in conjunction with transdermal oestradiol, on the cardiovascular system, lipid profile and coagulation cascade (174). There is a double-blind randomised controlled trial recruiting in America comparing DHEA versus placebo on pregnancy rates in POF (114). However, given that bone health is a major concern for women with POF and the fact that osteoporosis and resultant fractures can have major implications in later life, further research needs to be carried out in this area. Sexual function is also a concern in POF and in cross-sectional studies has been consistently found to be lower than controls. Our finding of a trend towards improved sexual function with HRT needs to be

confirmed in larger studies and the role of testosterone replacement in spontaneous POF, as opposed to surgically menopausal women, also needs to be evaluated.

Chapter 7 Appendices

7.1 Appendix A: Inclusion and Exclusion Criteria

Inclusion criteria

Age 18-44 years

Women with a diagnosis of POF within the last 36 months (with documented FSH level >30IU on 2 occasions 4-8 weeks apart)

Ability to understand English

Written informed consent for participation in the trial

Exclusion criteria

Age less than 18 or over 44 years

Current desire for pregnancy is an exclusion criterion from the active treatment group (as she may be randomised to take COCP). However, she could elect to take part in the no treatment group.

Women with absolute contraindications to hormone treatment will be excluded from the active treatment group (i.e. personal history of thromboembolic disease, oestrogen dependent malignancies, and personal history of focal migraine).

Women taking medication for high cholesterol or found to have raised cholesterol levels requiring treatment on initial assessment.

Untreated thyroid disease

7.2 Appendix B: Modified Greene Climacteric Scale

Modified Greene Climacteric Scale

Please indicate the extent to which you are bothered at the moment by any of these symptoms by ringing the appropriate number.

SEVERITY OF PROBLEM IS SCORED AS FOLLOWS:

SCORE
0……. None
1…….Mild
2…….Moderate
3…….Severe

Heart beating quickly and strongly	0	1	2	3		Feeling dizzy or faint	0	1	2	3
Feeling tense or nervous	0	1	2	3		Pressure or tightness in head or body	0	1	2	3
Difficulty in sleeping	0	1	2	3		Parts of body feeling numb or tingling	0	1	2	3
Excitable	0	1	2	3		Headaches	0	1	2	3
Attacks of panic	0	1	2	3		Muscle or joint pains	0	1	2	3
Difficulty in concentrating	0	1	2	3		Loss of feeling in hands or feet	0	1	2	3
Feeling tired or lacking in energy	0	1	2	3		Breathing difficulties	0	1	2	3
Loss of interest in most things	0	1	2	3		Hot flushes	0	1	2	3
Feeling unhappy or depressed	0	1	2	3		Sweating at night	0	1	2	3
Crying spells	0	1	2	3		Loss of interest in sex	0	1	2	3
Irritability	0	1	2	3						

7.3 Appendix C: Menopause Symptoms-Treatment Satisfaction Questionnaire

Menopause Symptoms Treatment Satisfaction Questionnaire

We are interested in learning what you think about the pill you've been taking as a part of this study. Please rate how satisfied you've been with each of the following during the past 4 weeks by checking one box for each question.

1. During the past 4 weeks, how satisfied have you been with the ability of the study medication to control your hot flashes during the day?

☐	☐	☐	☐	☐
Extremely Satisfied	Satisfied	Neutral	Dissatisfied	Extremely Dissatisfied

2. During the past 4 weeks, how satisfied have you been with the ability of the study medication to control your hot flashes during the night?

☐	☐	☐	☐	☐
Extremely Satisfied	Satisfied	Neutral	Dissatisfied	Extremely Dissatisfied

3. During the past 4 weeks, how satisfied have you been with the effect of the study medication on the quality of your sleep?

☐	☐	☐	☐	☐
Extremely Satisfied	Satisfied	Neutral	Dissatisfied	Extremely Dissatisfied

4. During the past 4 weeks, how satisfied have you been with the effect of the study medication on your mood or emotions?

☐	☐	☐	☐	☐
Extremely Satisfied	Satisfied	Neutral	Dissatisfied	Extremely Dissatisfied

5. During the past 4 weeks, how satisfied have you been with the effect of the study medication on your interest in sex?

☐	☐	☐	☐	☐
Extremely Satisfied	Satisfied	Neutral	Dissatisfied	Extremely Dissatisfied

6. During the past 4 weeks, how satisfied have you been with the effect of the study medication on your ability to concentrate?

☐	☐	☐	☐	☐
Extremely Satisfied	Satisfied	Neutral	Dissatisfied	Extremely Dissatisfied

7. While taking some medications, some people may experience side effects. How satisfied have you been with the tolerability (lack of bothersome side effects) of the study medication during the past 4 weeks?

☐	☐	☐	☐	☐
Extremely Satisfied	Satisfied	Neutral	Dissatisfied	Extremely Dissatisfied

8. During the past 4 weeks, overall, how satisfied have you been with the study medication?

☐	☐	☐	☐	☐
Extremely Satisfied	Satisfied	Neutral	Dissatisfied	Extremely Dissatisfied

7.4 Appendix D: Brief Profile of Female Sexual Function

Brief Profile of Female Sexual Function Questionnaire

In this questionnaire you will be asked about your sexual feelings and activity. Read each statement carefully and select the response that best corresponds to your experience over the past 2-3 months

I felt like having sex

Never Seldom Sometimes Often Very often Always

I was unhappy about my lack of interest in sex

Never Seldom Sometimes Often Very often Always

Getting aroused took forever

Never Seldom Sometimes Often Very often Always

I felt sexually numb

Never Seldom Sometimes Often Very often Always

I lacked sexual desire

Never Seldom Sometimes Often Very often Always

I felt disappointed by my lack of interest in sex

Never Seldom Sometimes Often Very often Always

I reached orgasm easily

Never Seldom Sometimes Often Very often Always

7.5 Appendix E: Patient-Health Questionnaire-9

Patient Health Questionnaire – 9

Over the last 2 weeks, how often have you been bothered by any of the following problems?

1. Little interest or pleasure in doing things

Not at all | Several days | More than half the days | Nearly every day

2. Feeling down, depressed, or hopeless

Not at all | Several days | More than half the days | Nearly every day

3. Trouble falling or staying asleep, or sleeping too much

Not at all | Several days | More than half the days | Nearly every day

4. Feeling tired or having little energy

Not at all | Several days | More than half the days | Nearly every day

5. Poor appetite or overeating

Not at all | Several days | More than half the days | Nearly every day

6. Feeling bad about yourself – or that you are a failure or have let yourself or your family down

Not at all | Several days | More than half the days | Nearly every day

7. Trouble concentrating on things, such as reading the newspaper or watching television

Not at all | Several days | More than half the days | Nearly every day

8. Moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual.

Not at all | Several days | More than half the days | Nearly every day

9. Thoughts that you would be better off dead or hurting yourself in some way

Not at all | Several days | More than half the days | Nearly every day

If you checked off <u>any</u> problems, how <u>difficult</u> have these problems made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all | Somewhat difficult | Very difficult | Extremely difficult

Chapter 8 References

1. Pitkin J, Rees MCP, Gray S, Lumsden MA, Marsden J, Stevenson JC, et al. Management of premature menopause. Menopause international. 2007;13(1):44-5.
2. Coulam CB, Adamson SC, Annegers JF. Incidence of premature ovarian failure. Obstetrics and gynecology. 1986;67(4):604-6.
3. van Kasteren YM, Schoemaker J. Premature ovarian failure: a systematic review on therapeutic interventions to restore ovarian function and achieve pregnancy. Human reproduction update. 1999;5(5):483-92.
4. Welt CK, Hall JE, Adams JM, Taylor AE. Relationship of estradiol and inhibin to the follicle-stimulating hormone variability in hypergonadotropic hypogonadism or premature ovarian failure. The Journal of clinical endocrinology and metabolism. 2005;90(2):826-30.
5. Goldacre MJ, Roberts SE, Yeates D. Mortality after admission to hospital with fractured neck of femur: database study. BMJ (Clinical research ed). 2002;325(7369):868-9.
6. Popat VB, Calis KA, Vanderhoof VH, Cizza G, Reynolds JC, Sebring N, et al. Bone mineral density in estrogen-deficient young women. J Clin Endocrinol Metab. 2009;94(7):2277-83.
7. Uygur D, Sengul O, Bayar D, Erdinc S, Batioglu S, Mollamahmutoglu L. Bone loss in young women with premature ovarian failure. Archives of gynecology and obstetrics. 2005;273(1):17-9.
8. Miller KK, Klibanski A. Clinical review 106: Amenorrheic bone loss. The Journal of clinical endocrinology and metabolism. 1999;84(6):1775-83.
9. Davies MC, Gulekli B, Jacobs HS. Osteoporosis in Turner's syndrome and other forms of primary amenorrhoea. Clinical Endocrinology. 1995;43(6):741-6.
10. van Der Voort DJM, van Der Weijer PHM, Barentsen R. Early menopause: increased fracture risk at older age. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2003;14(6):525-30.
11. Atsma F, Bartelink M-LEL, Grobbee DE, van der Schouw YT. Postmenopausal status and early menopause as independent risk factors for cardiovascular disease: a meta-analysis. Menopause (New York, N Y). 2006;13(2):265-79.
12. van der Schouw YT, van der Graaf Y, Steyerberg EW, Eijkemans JC, Banga JD. Age at menopause as a risk factor for cardiovascular mortality. Lancet. 1996;347(9003):714-8.
13. Knauff EAH, Westerveld HE, Goverde AJ, Eijkemans MJ, Valkenburg O, van Santbrink EJP, et al. Lipid profile of women with premature ovarian failure. Menopause (New York, N Y). 2008;15(5):919-23.
14. Liao KL, Wood N, Conway GS. Premature menopause and psychological well-being. Journal of psychosomatic obstetrics and gynaecology. 2000;21(3):167-74.
15. van der Stege JG, Groen H, van Zadelhoff SJN, Lambalk CB, Braat DDM, van Kasteren YM, et al. Decreased androgen concentrations and diminished general and sexual well-being in women with premature ovarian failure. Menopause (New York, N Y). 2008;15(1):23-31.
16. Kalu E, Panay N. Spontaneous premature ovarian failure: management challenges. Gynecological endocrinology : the official journal of the International Society of Gynecological Endocrinology. 2008;24(5):273-9.
17. Nelson LM, Covington SN, Rebar RW. An update: spontaneous premature ovarian failure is not an early menopause. Fertility and sterility. 2005;83(5):1327-32.
18. McKinlay SM. The normal menopause transition: an overview. Maturitas. 1996;23(2):137-45.
19. Faddy MJ. Follicle dynamics during ovarian ageing. Molecular and cellular endocrinology. 2000;163(1-2):43-8.
20. Hansen KR, Knowlton NS, Thyer AC, Charleston JS, Soules MR, Klein NA. A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. Hum Reprod. 2008;23(3):699-708.

21. Murray A, Webb J, Grimley S, Conway G, Jacobs P. Studies of FRAXA and FRAXE in women with premature ovarian failure. Journal of Medical Genetics. 1998;35(8):637-40.
22. Conway GS, Kaltsas G, Patel A, Davies MC, Jacobs HS. Characterization of idiopathic premature ovarian failure. Fertility & Sterility. 1996;65(2):337-41.
23. Cartwright B, Holloway D, Grace J, Robinson J, Rymer J. A service evaluation of women attending the menopause/premature ovarian failure clinic of a tertiary referral centre. J Obstet Gynaecol. 2012;32(4):357-61.
24. Bachelot A, Rouxel A, Massin N, Dulon J, Courtillot C, Matuchansky C, et al. Phenotyping and genetic studies of 357 consecutive patients presenting with premature ovarian failure. European Journal of Endocrinology. 2009;161(1):179-87.
25. van Kasteren YM, Hundscheid RD, Smits AP, Cremers FP, van Zonneveld P, Braat DD. Familial idiopathic premature ovarian failure: an overrated and underestimated genetic disease? Hum Reprod. 1999;14(10):2455-9.
26. Vujovic S. Aetiology of premature ovarian failure. Menopause Int. 2009;15(2):72-5.
27. Sherman SL. Premature ovarian failure in the fragile X syndrome. Am J Med Genet. 2000;97(3):189-94.
28. Corrigan EC, Raygada MJ, Vanderhoof VH, Nelson LM. A woman with spontaneous premature ovarian failure gives birth to a child with fragile X syndrome. Fertil Steril. 2005;84(5):1508.
29. Hagerman RJ, Leavitt BR, Farzin F, Jacquemont S, Greco CM, Brunberg JA, et al. Fragile-X-associated tremor/ataxia syndrome (FXTAS) in females with the FMR1 premutation. Am J Hum Genet. 2004;74(5):1051-6.
30. Pesso R, Berkenstadt M, Cuckle H, Gak E, Peleg L, Frydman M, et al. Screening for fragile X syndrome in women of reproductive age. Prenat Diagn. 2000;20(8):611-4.
31. Hagerman RJ, Hagerman PJ. The fragile X premutation: into the phenotypic fold. Curr Opin Genet Dev. 2002;12(3):278-83.
32. Toledano-Alhadef H, Basel-Vanagaite L, Magal N, Davidov B, Ehrlich S, Drasinover V, et al. Fragile-X carrier screening and the prevalence of premutation and full-mutation carriers in Israel. Am J Hum Genet. 2001;69(2):351-60.
33. Wittenberger MD, Hagerman RJ, Sherman SL, McConkie-Rosell A, Welt CK, Rebar RW, et al. The FMR1 premutation and reproduction. Fertil Steril. 2007;87(3):456-65.
34. Sybert VP, McCauley E. Turner's syndrome. N Engl J Med. 2004;351(12):1227-38.
35. Mortensen KH, Rohde MD, Uldbjerg N, Gravholt CH. Repeated spontaneous pregnancies in 45,X Turner syndrome. Obstetrics & Gynecology. 2010;115(2 Pt 2):446-9.
36. Oktay K, Rodriguez-Wallberg KA, Sahin G. Fertility preservation by ovarian stimulation and oocyte cryopreservation in a 14-year-old adolescent with Turner syndrome mosaicism and impending premature ovarian failure. Fertility & Sterility. 2010;94(2):753.e15-9.
37. Lakhal B, Braham R, Berguigua R, Bouali N, Zaouali M, Chaieb M, et al. Cytogenetic analyses of premature ovarian failure using karyotyping and interphase fluorescence in situ hybridization (FISH) in a group of 1000 patients. Clinical Genetics. 2010;78(2):181-5.
38. Rees M, Stevenson J, Hope S, Rozenberg S, Palacios S. Management of the Menopause. 5th Edition ed: Royal Society of Medicine Press Ltd and British Menopause Society Publications Ltd; 2009.
39. Persani L, Rossetti R, Cacciatore C, Bonomi M. Primary Ovarian Insufficiency: X chromosome defects and autoimmunity. Journal of Autoimmunity. 2009;33(1):35-41.
40. Laml T, Preyer O, Umek W, Hengstschlager M, Hanzal H. Genetic disorders in premature ovarian failure. Human Reproduction Update. 2002;8(5):483-91.
41. Chand AL, Harrison CA, Shelling AN. Inhibin and premature ovarian failure. Human Reproduction Update. 2010;16(1):39-50.
42. Luborsky J, Llanes B, Davies S, Binor Z, Radwanska E, Pong R. Ovarian autoimmunity: greater frequency of autoantibodies in premature menopause and unexplained infertility than in the general population. Clinical Immunology. 1999;90(3):368-74.
43. Goswami R, Marwaha RK, Goswami D, Gupta N, Ray D, Tomar N, et al. Prevalence of thyroid autoimmunity in sporadic idiopathic hypoparathyroidism in comparison to type 1

diabetes and premature ovarian failure. Journal of Clinical Endocrinology & Metabolism. 2006;91(11):4256-9.
44. Bakalov VK, Vanderhoof VH, Bondy CA, Nelson LM. Adrenal antibodies detect asymptomatic auto-immune adrenal insufficiency in young women with spontaneous premature ovarian failure. Human Reproduction. 2002;17(8):2096-100.
45. Betterle C, Coco G, Zanchetta R. Adrenal cortex autoantibodies in subjects with normal adrenal function. Best Practice & Research Clinical Endocrinology & Metabolism. 2005;19(1):85-99.
46. Faddy MJ, Gosden RG, Gougeon A, Richardson SJ, Nelson JF. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. Hum Reprod. 1992;7(10):1342-6.
47. Richardson SJ, Senikas V, Nelson JF. Follicular depletion during the menopausal transition: evidence for accelerated loss and ultimate exhaustion. The Journal of clinical endocrinology and metabolism. 1987;65(6):1231-7.
48. Schmidt KT, Larsen EC, Andersen CY, Andersen AN. Risk of ovarian failure and fertility preserving methods in girls and adolescents with a malignant disease. BJOG: An International Journal of Obstetrics & Gynaecology. 2010;117(2):163-74.
49. Morrison JC, Givens JR, Wiser WL, Fish SA. Mumps oophoritis: a cause of premature menopause. Fertility & Sterility. 1975;26(7):655-9.
50. Ohl J, Partisani M, Demangeat C, Binder-Foucard F, Nisand I, Lang JM. [Alterations of ovarian reserve tests in Human Immunodeficiency Virus (HIV)-infected women]. Gynecologie, Obstetrique & Fertilite. 2010;38(5):313-7.
51. Benaglia L, Somigliana E, Vighi V, Ragni G, Vercellini P, Fedele L. Rate of severe ovarian damage following surgery for endometriomas. Human Reproduction. 2010;25(3):678-82.
52. Islam RC, R. The impact of premature ovarian failure on quality of life: Results from the UK 1958 Birth Cohort. Human Reproduction2011. p. i108.
53. Maclennan AH, Broadbent JL, Lester S, Moore V. Oral oestrogen and combined oestrogen/progestogen therapy versus placebo for hot flushes. Cochrane database of systematic reviews (Online). 2004(4):CD002978.
54. Schmidt PJ, Luff JA, Haq NA, Vanderhoof VH, Koziol DE, Calis KA, et al. Depression in women with spontaneous 46, XX primary ovarian insufficiency. Journal of Clinical Endocrinology & Metabolism. 2011;96(2):E278-87.
55. Singer D, Mann E, Hunter MS, Pitkin J, Panay N. The silent grief: psychosocial aspects of premature ovarian failure. Climacteric. 2011;14(4):428-37.
56. Absolom K, Eiser C, Turner L, Ledger W, Ross R, Davies H, et al. Ovarian failure following cancer treatment: current management and quality of life. Hum Reprod. 2008;23(11):2506-12.
57. Piccioni P, Scirpa P, D'Emilio I, Sora F, Scarciglia M, Laurenti L, et al. Hormonal replacement therapy after stem cell transplantation. Maturitas. 2004;49(4):327-33.
58. Davis M, Ventura JL, Wieners M, Covington SN, Vanderhoof VH, Ryan ME, et al. The psychosocial transition associated with spontaneous 46,XX primary ovarian insufficiency: illness uncertainty, stigma, goal flexibility, and purpose in life as factors in emotional health. Fertility & Sterility. 2010;93(7):2321-9.
59. Kalantaridou SN, Vanderhoof VH, Calis KA, Corrigan EC, Troendle JF, Nelson LM. Sexual function in young women with spontaneous 46,XX primary ovarian insufficiency. Fertility and sterility. 2008;90(5):1805-11.
60. Davis SR. Sexual health consequences of premature ovarian failure. Menopause. 2008;15(1):5-6.
61. Janse F, Tanahatoe SJ, Eijkemans MJ, Fauser BC. Testosterone concentrations, using different assays, in different types of ovarian insufficiency: a systematic review and meta-analysis. Hum Reprod Update. 2012;18(4):405-19.
62. de Almeida DMB, Benetti-Pinto CL, Makuch MY. Sexual function of women with premature ovarian failure. Menopause. 2011;18(3):262-6.

63. WHO. Prevention and Management of Osteoporosis. Geneva: World Health Organisation, 2003 Contract No.: 921.
64. Blake GM, Fogelman I. The role of DXA bone density scans in the diagnosis and treatment of osteoporosis. Postgraduate medical journal. 2007;83(982):509-17.
65. Johnell O, Kanis JA, Oden A, Johansson H, De Laet C, Delmas P, et al. Predictive value of BMD for hip and other fractures. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2005;20(7):1185-94.
66. Lips P, van Schoor NM. Quality of life in patients with osteoporosis. Osteoporos Int. 2005;16(5):447-55.
67. Adachi JD, Adami S, Gehlbach S, Anderson FA, Jr., Boonen S, Chapurlat RD, et al. Impact of prevalent fractures on quality of life: baseline results from the global longitudinal study of osteoporosis in women. Mayo Clin Proc. 2010;85(9):806-13.
68. Abrahamsen B, van Staa T, Ariely R, Olson M, Cooper C. Excess mortality following hip fracture: a systematic epidemiological review. Osteoporos Int. 2009;20(10):1633-50.
69. Siris ES, Chen YT, Abbott TA, Barrett-Connor E, Miller PD, Wehren LE, et al. Bone mineral density thresholds for pharmacological intervention to prevent fractures. Arch Intern Med. 2004;164(10):1108-12.
70. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2008;19(4):385-97.
71. Poole KE, Compston JE. Osteoporosis and its management. BMJ. 2006;333(7581):1251-6.
72. Raggatt LJ, Partridge NC. Cellular and molecular mechanisms of bone remodeling. J Biol Chem. 2010;285(33):25103-8.
73. Sipos W, Pietschmann P, Rauner M, Kerschan-Schindl K, Patsch J. Pathophysiology of osteoporosis. Wiener medizinische Wochenschrift (1946). 2009;159(9-10):230-4.
74. Khosla S. Estrogen and bone: insights from estrogen-resistant, aromatase-deficient, and normal men. Bone. 2008;43(3):414-7.
75. Lufkin EG, Wahner HW, O'Fallon WM, Hodgson SF, Kotowicz MA, Lane AW, et al. Treatment of postmenopausal osteoporosis with transdermal estrogen. Annals of Internal Medicine. 1992;117(1):1-9.
76. Charatcharoenwitthaya N, Khosla S, Atkinson EJ, McCready LK, Riggs BL. Effect of blockade of TNF-alpha and interleukin-1 action on bone resorption in early postmenopausal women. J Bone Miner Res. 2007;22(5):724-9.
77. Wells G, Tugwell P, Shea B, Guyatt G, Peterson J, Zytaruk N, et al. Meta-analyses of therapies for postmenopausal osteoporosis. V. Meta-analysis of the efficacy of hormone replacement therapy in treating and preventing osteoporosis in postmenopausal women. Endocrine reviews. 2002;23(4):529-39.
78. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. Jama. 2002;288(3):321-33.
79. Gambacciani M, Ciaponi M, Cappagli B, Benussi C, Genazzani AR. Longitudinal evaluation of perimenopausal femoral bone loss: effects of a low-dose oral contraceptive preparation on bone mineral density and metabolism. Osteoporosis International. 2000;11(6):544-8.
80. Volpe A, Amram A, Cagnacci A, Battaglia C. Biochemical aspects of hormonal contraception: effects on bone metabolism. European Journal of Contraception & Reproductive Health Care. 1997;2(2):123-6.
81. Jones KP, Ravnikar VA, Tulchinsky D, Schiff I. Comparison of bone density in amenorrheic women due to athletics, weight loss, and premature menopause. Obstetrics and gynecology. 1985;66(1):5-8.

82. Svejme O, Ahlborg HG, Nilsson J, Karlsson MK. Early menopause and risk of osteoporosis, fracture and mortality: a 34-year prospective observational study in 390 women. BJOG. 2012;119(7):810-6.
83. Crofton PM, Evans N, Bath LE, Warner P, Whitehead TJ, Critchley HOD, et al. Physiological versus standard sex steroid replacement in young women with premature ovarian failure: effects on bone mass acquisition and turnover. Clinical Endocrinology. 2010;73(6):707-14.
84. Fink E, Cormier C, Steinmetz P, Kindermans C, Le Bouc Y, Souberbielle JC. Differences in the capacity of several biochemical bone markers to assess high bone turnover in early menopause and response to alendronate therapy. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2000;11(4):295-303.
85. Delmas PD, Eastell R, Garnero P, Seibel MJ, Stepan J, Committee of Scientific Advisors of the International Osteoporosis F. The use of biochemical markers of bone turnover in osteoporosis. Committee of Scientific Advisors of the International Osteoporosis Foundation. Osteoporosis International. 2000;11 Suppl 6:S2-17.
86. Garnero P, Sornay-Rendu E, Delmas PD. Decreased bone turnover in oral contraceptive users. Bone. 1995;16(5):499-503.
87. Guttmann H, Weiner Z, Nikolski E, Ish-Shalom S, Itskovitz-Eldor J, Aviram M, et al. Choosing an oestrogen replacement therapy in young adult women with Turner syndrome. Clinical endocrinology. 2001;54(2):159-64.
88. Grodstein F, Manson JE, Stampfer MJ. Hormone therapy and coronary heart disease: the role of time since menopause and age at hormone initiation.[see comment]. Journal of Women's Health. 2006;15(1):35-44.
89. Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. JAMA. 1998;280(7):605-13.
90. Vickers MR, MacLennan AH, Lawton B, Ford D, Martin J, Meredith SK, et al. Main morbidities recorded in the women's international study of long duration oestrogen after menopause (WISDOM): a randomised controlled trial of hormone replacement therapy in postmenopausal women. BMJ. 2007;335(7613):239.
91. Salpeter SR, Walsh JM, Greyber E, Salpeter EE. Brief report: Coronary heart disease events associated with hormone therapy in younger and older women. A meta-analysis. J Gen Intern Med. 2006;21(4):363-6.
92. Salpeter SR, Cheng J, Thabane L, Buckley NS, Salpeter EE. Bayesian meta-analysis of hormone therapy and mortality in younger postmenopausal women. Am J Med. 2009;122(11):1016-22 e1.
93. Rossouw JE, Prentice RL, Manson JE, Wu L, Barad D, Barnabei VM, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. JAMA : the journal of the American Medical Association. 2007;297(13):1465-77.
94. Schierbeck LL, Rejnmark L, Tofteng CL, Stilgren L, Eiken P, Mosekilde L, et al. Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women: randomised trial. BMJ. 2012;345:e6409.
95. Teede HJ. Sex hormones and the cardiovascular system: effects on arterial function in women. Clinical and experimental pharmacology & physiology. 2007;34(7):672-6.
96. Løkkegaard E, Jovanovic Z, Heitmann BL, Keiding N, Ottesen B, Pedersen AT. The association between early menopause and risk of ischaemic heart disease: influence of Hormone Therapy. Maturitas. 2006;53(2):226-33.
97. Rees M, Stevenson J. Primary prevention of coronary heart disease in women. Menopause international. 2008;14(1):40-5.
98. Kalantaridou SN, Naka KK, Papanikolaou E, Kazakos N, Kravariti M, Calis KA, et al. Impaired endothelial function in young women with premature ovarian failure: normalization

with hormone therapy. The Journal of clinical endocrinology and metabolism. 2004;89(8):3907-13.
99. Yorgun H, Tokgozoglu L, Canpolat U, Gurses KM, Bozdag G, Yapici Z, et al. The cardiovascular effects of premature ovarian failure. Int J Cardiol. 2013;168(1):506-10.
100. Goldmeier S, De Angelis K, Rabello Casali K, Vilodre C, Consolim-Colombo F, Bello Klein A, et al. Cardiovascular autonomic dysfunction in primary ovarian insufficiency: clinical and experimental evidence. Am J Transl Res. 2013;6(1):91-101.
101. Langrish JP, Mills NL, Bath LE, Warner P, Webb DJ, Kelnar CJ, et al. Cardiovascular effects of physiological and standard sex steroid replacement regimens in premature ovarian failure. Hypertension. 2009;53(5):805-11.
102. Stevenson JC, Crook D, Godsland IF. Influence of age and menopause on serum lipids and lipoproteins in healthy women. Atherosclerosis. 1993;98(1):83-90.
103. Ridker PM. C-reactive protein, inflammation, and cardiovascular disease: clinical update. Texas Heart Institute Journal. 2005;32(3):384-6.
104. Ridker PM, Hennekens CH, Buring JE, Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. New England Journal of Medicine. 2000;342(12):836-43.
105. Kurtz EG, Ridker PM, Rose LM, Cook NR, Everett BM, Buring JE, et al. Oral postmenopausal hormone therapy, C-reactive protein, and cardiovascular outcomes. Menopause. 2011;18(1):23-9.
106. Dardzińska JA, Rachoń D, Kuligowska-Jakubowska M, Aleksandrowicz-Wrona E, Płoszyński A, Wyrzykowski B, et al. Effects of metformin or an oral contraceptive containing cyproterone acetate on serum c-reactive protein, interleukin-6 and soluble vascular cell adhesion molecule-1 concentrations in women with polycystic ovary syndrome. Exp Clin Endocrinol Diabetes. 2014;122(2):118-25.
107. Stuenkel CA. Hold, raise, or fold: C-reactive protein, hormone therapy, and cardiovascular risk. Menopause. 2011;18(1):8-10.
108. Patel D, Jhamnani S, Ahmad S, Silverman A, Lindsay J. Discordant association of C-reactive protein with clinical events and coronary luminal narrowing in postmenopausal women: data from the Women's Angiographic Vitamin and Estrogen (WAVE) study. Clin Cardiol. 2013;36(9):535-41.
109. Gaskins AJ, Wilchesky M, Mumford SL, Whitcomb BW, Browne RW, Wactawski-Wende J, et al. Endogenous reproductive hormones and C-reactive protein across the menstrual cycle: the BioCycle Study. Am J Epidemiol. 2012;175(5):423-31.
110. Alzubaidi NH, Chapin HL, Vanderhoof VH, Calis KA, Nelson LM. Meeting the needs of young women with secondary amenorrhea and spontaneous premature ovarian failure. Obstetrics and gynecology. 2002;99(5 Pt 1):720-5.
111. Bidet M, Bachelot A, Touraine P. Premature ovarian failure: predictability of intermittent ovarian function and response to ovulation induction agents. Curr Opin Obstet Gynecol. 2008;20(4):416-20.
112. Mamas L, Mamas E. Dehydroepiandrosterone supplementation in assisted reproduction: rationale and results. Current Opinion in Obstetrics & Gynecology. 2009;21(4):306-8.
113. Mamas L, Mamas E. Premature ovarian failure and dehydroepiandrosterone. Fertility & Sterility. 2009;91(2):644-6.
114. http://www.controlled-trials.com/mrct/trial/602393/premature+ovarian+failure.
115. Oyesanya OA, Olufowobi O, Ross W, Sharif K, Afnan M. Prognosis of oocyte donation cycles: a prospective comparison of the in vitro fertilization-embryo transfer cycles of recipients who used shared oocytes versus those who used altruistic donors. Fertility & Sterility. 2009;92(3):930-6.
116. Ameratunga D, Weston G, Osianlis T, Catt J, Vollenhoven B. In vitro fertilisation (IVF) with donor eggs in post-menopausal women: are there differences in pregnancy outcomes in women with premature ovarian failure (POF) compared with women with physiological age-related menopause? Journal of Assisted Reproduction & Genetics. 2009;26(9-10):511-4.

117. Human Fertility and Embryology Association
118. Donnez J, Jadoul P, Squifflet J, Van Langendonckt A, Donnez O, Van Eyck A-S, et al. Ovarian tissue cryopreservation and transplantation in cancer patients. Best Practice & Research in Clinical Obstetrics & Gynaecology. 2010;24(1):87-100.
119. Henrich JB, Hughes JP, Kaufman SC, Brody DJ, Curtin LR. Limitations of follicle-stimulating hormone in assessing menopause status: findings from the National Health and Nutrition Examination Survey (NHANES 1999-2000)*. Menopause (New York, N Y). 2006;13(2):171-7.
120. Knauff EAH, Eijkemans MJC, Lambalk CB, ten Kate-Booij MJ, Hoek A, Beerendonk CCM, et al. Anti-Mullerian hormone, inhibin B, and antral follicle count in young women with ovarian failure. The Journal of clinical endocrinology and metabolism. 2009;94(3):786-92.
121. Khairy M, Clough A, El-Toukhy T, Coomarasamy A, Khalaf Y. Antral follicle count at down-regulation and prediction of poor ovarian response. Reproductive biomedicine online. 2008;17(4):508-14.
122. Scheffer GJ, Broekmans FJ, Dorland M, Habbema JD, Looman CW, te Velde ER. Antral follicle counts by transvaginal ultrasonography are related to age in women with proven natural fertility. Fertility and sterility. 1999;72(5):845-51.
123. Pache TD, Wladimiroff JW, de Jong FH, Hop WC, Fauser BC. Growth patterns of nondominant ovarian follicles during the normal menstrual cycle. Fertility and sterility. 1990;54(4):638-42.
124. Pavlik EJ, DePriest PD, Gallion HH, Ueland FR, Reedy MB, Kryscio RJ, et al. Ovarian volume related to age. Gynecologic oncology. 2000;77(3):410-2.
125. Hendriks DJ, Kwee J, Mol BWJ, te Velde ER, Broekmans FJM. Ultrasonography as a tool for the prediction of outcome in IVF patients: a comparative meta-analysis of ovarian volume and antral follicle count. Fertility and sterility. 2007;87(4):764-75.
126. de Carvalho BR, Rosa e Silva ACJdS, Rosa e Silva JC, dos Reis RM, Ferriani RA, Silva de Sa MF. Ovarian reserve evaluation: state of the art. Journal of assisted reproduction and genetics. 2008;25(7):311-22.
127. Weenen C, Laven JSE, Von Bergh ARM, Cranfield M, Groome NP, Visser JA, et al. Anti-Mullerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. Molecular human reproduction. 2004;10(2):77-83.
128. Visser JA, Themmen APN. Anti-Mullerian hormone and folliculogenesis. Molecular & Cellular Endocrinology. 2005;234(1-2):81-6.
129. Broer SL, Mol BWJ, Hendriks D, Broekmans FJM. The role of antimullerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. Fertility & Sterility. 2009;91(3):705-14.
130. van Disseldorp J, Faddy MJ, Themmen AP, de Jong FH, Peeters PH, van der Schouw YT, et al. Relationship of serum antimullerian hormone concentration to age at menopause. J Clin Endocrinol Metab. 2008;93(6):2129-34.
131. Lambalk CB, van Disseldorp J, de Koning CH, Broekmans FJ. Testing ovarian reserve to predict age at menopause. Maturitas. 2009;63(4):280-91.
132. Broekmans FJ, Kwee J, Hendriks DJ, Mol BW, Lambalk CB. A systematic review of tests predicting ovarian reserve and IVF outcome. Human reproduction update. 2006;12(6):685-718.
133. La Marca A, Pati M, Orvieto R, Stabile G, Carducci Artenisio A, Volpe A. Serum anti-mullerian hormone levels in women with secondary amenorrhea. Fertility & Sterility. 2006;85(5):1547-9.
134. Massin N, Meduri G, Bachelot A, Misrahi M, Kuttenn F, Touraine P. Evaluation of different markers of the ovarian reserve in patients presenting with premature ovarian failure. Molecular & Cellular Endocrinology. 2008;282(1-2):95-100.
135. Mittal M SM, Narvekar N, Panay N, Hamoda H. A cross-sectional national survey assessing the clinical attitudes to the management of premature ovarian insufficiency. 23rd Annual Conference of the British Menopause Society; Stratford Upon Avon: Menopause International; 2013. p. 108.

136. Opinions survey report no 41. Contraception and Sexual Health Office for National Statistics, 2008-9.
137. Hanton L, Axelrod L, Bakalov V, Bondy CA. The importance of estrogen replacement in young women with Turner syndrome. J Womens Health (Larchmt). 2003;12(10):971-7.
138. Ng C, Sacks G, Turner C, Reddy N, Panay N. Impact of alarmist media on HRT use for premature menopause. ESHRE; Berlin. Abstracts of the 20th Annual Meeting of the ESHRE, Berlin, Germany, 27–30 June 20042004.

139. Mashchak CA, Lobo RA, Dozono-Takano R, Eggena P, Nakamura RM, Brenner PF, et al. Comparison of pharmacodynamic properties of various estrogen formulations. Am J Obstet Gynecol. 1982;144(5):511-8.

140. Scottish Intercollegiate Guidelines Network (SIGN) Management of Osteoporosis SIGN publication number 71. 2003.
141. Blake GM, Fogelman I. Technical principles of dual energy x-ray absorptiometry. Seminars in nuclear medicine. 1997;27(3):210-28.
142. Patel R, Blake GM, Rymer J, Fogelman I. Long-term precision of DXA scanning assessed over seven years in forty postmenopausal women. Osteoporos Int. 2000;11(1):68-75.
143. Blake GM, Fogelman I. An update on dual-energy x-ray absorptiometry. Semin Nucl Med. 2010;40(1):62-73.
144. Noon E, Singh S, Cuzick J, Spector TD, Williams FM, Frost ML, et al. Significant differences in UK and US female bone density reference ranges. Osteoporos Int. 2010;21(11):1871-80.
145. Garnero P, Vergnaud P, Hoyle N. Evaluation of a fully automated serum assay for total N-terminal propeptide of type I collagen in postmenopausal osteoporosis. Clin Chem. 2008;54(1):188-96.
146. Garnero P, Borel O, Delmas PD. Evaluation of a fully automated serum assay for C-terminal cross-linking telopeptide of type I collagen in osteoporosis. Clin Chem. 2001;47(4):694-702.
147. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clinical chemistry. 1972;18(6):499-502.
148. Wallace AM, Faye SA, Fleming R, Nelson SM. A multicentre evaluation of the new Beckman Coulter anti-Mullerian hormone immunoassay (AMH Gen II). Ann Clin Biochem. 2011;48(Pt 4):370-3.
149. Scheffer GJ, Broekmans FJ, Bancsi LF, Habbema JD, Looman CW, Te Velde ER. Quantitative transvaginal two- and three-dimensional sonography of the ovaries: reproducibility of antral follicle counts. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2002;20(3):270-5.
150. Greene JG. Constructing a standard climacteric scale. Maturitas. 2008;61(1-2):78-84.
151. Vasconcelos-Raposo J, Coelho E, Fernandes HM, Rodrigues C, Moreira H, Teixeira C. Factor structure and normative data of the Greene Climacteric Scale among postmenopausal Portuguese women. Maturitas. 2012;72(3):256-62.
152. Buckler H, Al-Azzawi F. The effect of a novel vaginal ring delivering oestradiol acetate on climacteric symptoms in postmenopausal women. BJOG. 2003;110(8):753-9.
153. Davis SR, Davison SL, Wilson S, Shepherd J, Lawton B. Intranasal versus transdermal matrix oestrogen replacement in Australasian women. Maturitas. 2005;51(2):163-71.
154. Hill CD, Fehnel SE, Bobula JD, Yu H, McLeod LD. Development and preliminary validation of the Menopause Symptoms Treatment Satisfaction Questionnaire (MS-TSQ). Menopause (New York, N Y). 2007;14(6):1047-55.
155. Wyrwich KW, Spratt DI, Gass M, Yu H, Bobula JD. Identifying meaningful differences in vasomotor symptoms among menopausal women. Menopause. 2008;15(4 Pt 1):698-705.

156. Rust J, Derogatis L, Rodenberg C, Koochaki P, Schmitt S, Golombok S. Development and validation of a new screening tool for hypoactive sexual desire disorder: The Brief Profile of Female Sexual Function (B-PFSF). Gynecol Endocrinol. 2007;23(11):638-44.
157. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. Journal of general internal medicine. 2001;16(9):606-13.
158. Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). Journal of affective disorders. 2004;81(1):61-6.
159. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. General hospital psychiatry. 2006;28(1):71-7.
160. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. Journal of general internal medicine. 2006;21(6):547-52.
161. Ware J E KM, Bjorner J B, Turner-Bowker D M, Gandek B, Maruish ME. User's Manual for the SF-36v2 Health Survey (2nd Ed). 2 ed. 2 ed. ed: QualityMetric Incorporated; 2008.
162. Sadatsafavi M, Moayyeri A, Wang L, Leslie WD. Optimal decision criterion for detecting change in bone mineral density during serial monitoring: a Bayesian approach. Osteoporosis International. 2008;19(11):1589-96.
163. Kaptoge S, Reid DM, Scheidt-Nave C, Poor G, Pols HAP, Khaw KT, et al. Geographic and other determinants of BMD change in European men and women at the hip and spine. a population-based study from the Network in Europe for Male Osteoporosis (NEMO). Bone. 2007;40(3):662-73.
164. Kenward MG, Carpenter J. Multiple imputation: current perspectives. Stat Methods Med Res. 2007;16(3):199-218.
165. Lopez LM, Grimes DA, Schulz KF, Curtis KM, Chen M. Steroidal contraceptives: effect on bone fractures in women. Cochrane Database Syst Rev. 2014;6:CD006033.
166. Herrmann M, Seibel MJ. The effects of hormonal contraceptives on bone turnover markers and bone health. Clin Endocrinol (Oxf). 2010;72(5):571-83.
167. Han X, McShane M, Sahertian R, White C, Ledger W. Pre-mixing serum samples with assay buffer is a prerequisite for reproducible anti-Mullerian hormone measurement using the Beckman Coulter Gen II assay. Hum Reprod. 2014;29(5):1042-8.
168. Ledger WL. Measurement of antimullerian hormone: not as straightforward as it seems. Fertil Steril. 2014;101(2):339.
169. Rees M, Stevenson J, Hope S, Rozenberg S, Palacios S. Management of the Menopause. 5th Edition ed: Royal Society of Medicine Press Ltd and British Menopause Society Publications Ltd; 2009.
170. Pickar JH, Thorneycroft I, Whitehead M. Effects of hormone replacement therapy on the endometrium and lipid parameters: a review of randomized clinical trials, 1985 to 1995. American journal of obstetrics and gynecology. 1998;178(5):1087-99.
171. Zimmerman Y, Eijkemans MJ, Coelingh Bennink HJ, Blankenstein MA, Fauser BC. The effect of combined oral contraception on testosterone levels in healthy women: a systematic review and meta-analysis. Hum Reprod Update. 2014;20(1):76-105.
172. Warnock JK, Clayton A, Croft H, Segraves R, Biggs FC. Comparison of androgens in women with hypoactive sexual desire disorder: those on combined oral contraceptives (COCs) vs. those not on COCs. J Sex Med. 2006;3(5):878-82.
173. Cumming GP, Currie HD, Moncur R, Lee AJ. Web-based survey on the effect of menopause on women's libido in a computer-literate population. Menopause Int. 2009;15(1):8-12.
174. Mittal M, Savvas M, Arya R, McEniery C, Narvekar N, Cardozo L, et al. A randomised controlled trial comparing the effects of micronized progesterone to medroxyprogesterone acetate on cardiovascular health, lipid metabolism and the coagulation cascade in women with premature ovarian insufficiency: study protocol and review of the literature. Menopause Int. 2013;19(3):127-32.